THE BETTER ANGELS OF OUR NATURE

AN EXPLORATION OF CONSTRUCTIVISM IN ETHICS, THE NORMATIVE DIMENSION OF PEAK EXPERIENCES, AND THE PHENOMENOLOGY OF MEDITATION PRACTICE

Maxwell Mitchell

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

Department of Philosophy

July 2023

ABSTRACT

2

This thesis is about constructivism in ethics. More specifically, it is about an unusual and little-known meta-ethical theory called 'Humean constructivism'. I try to do three main things in this thesis. First, I try to illuminate a constellation of positions in logical space: Humean constructivism and its constructivist neighbours. Second, I try to paint a nuanced picture of the strengths and weaknesses of the view – I do this in the hope that the resulting portrait, when viewed from some distance, will appear attractive and interesting to potential buyers. Third, I try to *enhance* Humean constructivism. This third project, that of enhancement, takes up the latter half of the thesis; it will involve a strange fusion of topics: meta-ethics, moral psychology, and the phenomenology of peak experiences and meditative experiences.

Ultimately, I hope to show that Humean constructivism is one of the most plausible forms of subjectivism on the market right now and is worth serious consideration as a theory of ethical truth. Let me add, however, that parts of this thesis may be of broader interest to those whose speciality is not in meta-ethics *per se* but in areas such as phenomenology and philosophy of mind. Chapter 4 of this thesis, for instance, involves a lengthy analysis of certain kinds of transcendent, numinous, perspective-altering experiences – a.k.a. *peak* experiences – and of the first-person, experiential side of various strands of meditation practice, such as *vipassanā*.

CONTENTS

The moving Moon went up the sky,
And no where did abide:
Softly she was going up,
And a star or two beside—

Her beams bemocked the sultry main,
Like April hoar-frost spread;
But where the ship's huge shadow lay,
The charmèd water burnt alway
A still and awful red.

Beyond the shadow of the ship,
I watched the water-snakes:
They moved in tracks of shining white,
And when they reared, the elfish light
Fell off in hoary flakes.

Within the shadow of the ship
I watched their rich attire:
Blue, glossy green, and velvet black,
They coiled and swam; and every track
Was a flash of golden fire.

O happy living things! no tongue
Their beauty might declare:
A spring of love gushed from my heart,
And I blessed them unaware:
Sure my kind saint took pity on me,
And I blessed them unaware.

— Samuel Taylor Coleridge, *The Rime of the Ancient Mariner*

And those simple gifts, like other objects equally trivial—bread, oil, wine, milk —had regained for him, by their use in such religious service, that poetic, and as it were moral significance, which surely belongs to the means of our daily life, could we but break through the veil of our familiarity with things by no means vulgar in themselves.

— Walter Pater, *Marius the Epicurean*

Any act can be a prayer, if done as well as we are able.

— George R.R. Martin, *A Storm of Swords*

INTRODUCTION

Let me first say a word about the structure of this thesis. This thesis is structured more like an anthology of essays than like an overarching narrative oriented towards the defence of a core set of claims. That is to say, the four chapters that make up this thesis are, to a large extent, free-standing and self-contained. They are like self-contained units that combine to form a larger, cohesive whole. This is not to say that there are *no* connecting threads running through the various chapters, but just that they hang quite loosely. This is especially true of the final chapter, Chapter 4, which is almost like a miniature thesis nested within the thesis.

Broadly speaking, this thesis is a sustained exploration and defence of a meta-ethical theory known as 'Humean constructivism'. Humean constructivism is a niche form of constructivism that originates with Sharon Street (2008; 2010). It is a theory about the conditions under which ascriptions of moral reasons are true. In brief, the theory says that ethical truths – i.e. truths about the moral reasons that a particular agent possesses – are *constructed* by agents as opposed to being *discovered* by them. Such truths are constructed using the raw materials that are an agent's substantive evaluations – i.e. states of taking certain things/options in the world to be good or bad, right or wrong, better or worse, worthy or worthless, required or optional, virtuous or sinful, honourable or dishonourable, valuable or disvaluable, etc. – together with any non-normative facts that have a bearing on those evaluations (Street 2010: 366–367).[1]

A big part of what I take myself to be doing in this thesis is illuminating a position in logical space. But one cannot understand the position that is Humean constructivism without understanding the broader family of constructivist views to which it belongs. So, I will examine the constellation of constructivist views that surround it. The analogy I would draw is as follows. In this thesis, I am like a cartographer trying to map out an underground cave system. The cave system stands for the network of views in ethics that fall under the heading of 'constructivism'. The main subterranean cave that I will be poking around in throughout this thesis is that which relates to Humean constructivism. It is a cave

---

[1] I will flesh out this bare-bones summary in Chapter 1. For now, I merely aim to give a sense of the overall thrust of the thesis.

that is relatively dark and unexplored. Its contours and crannies are mostly uncharted; they are blank spaces on the map. But I cannot access this cave without first taking some detours through neighbouring caves. These include the caves relating to constructivism as a substantive normative position (as opposed to a *meta*-normative one). And importantly, they also include the cave that is Kantian constructivism.

Kantian constructivism is Humean constructivism's meta-ethical cousin (and arch-nemesis). There are major structural similarities between the two theories but also some key differences. Kantian constructivism finds its canonical expression in the work of Christine Korsgaard (1996). As I will show in Chapter 1, §1.2, Humean constructivism is largely defined in contrast and in opposition to Kantian constructivism (Street 2012: 45). As such, an important part of my exposition of Humean constructivism will involve comparing and contrasting it with Kantian constructivism and then offering some reasons for rejecting the latter position in favour of the former.

Just for some context, when it comes to the academic literature on meta-ethical constructivism, the Humean form has received far *less* attention than its Kantian cousin. If I had to assign a numerical figure to it, I would say that about five per cent of the relevant academic literature concerns the former position, with the remaining ninety-five per cent concentrated on the latter. To return to my underground cave analogy, Humean constructivism is a cave that remains largely unexplored and uncharted – which is part of the reason why I am interested in it. Kantian constructivism, by contrast, is a cave that has been well-traversed, and few mysteries remain hidden in its depths.

In addition to illuminating a position in logical space, I will also try to defend that position. My approach in this regard will mirror that of Sobel (2016: 17–18) in his defence of meta-ethical subjectivism. That is, I will not attempt to isolate a single consideration that is so powerful that it secures the truth of Humean constructivism regardless of its other costs. There *is* no single argument that conclusively establishes the truth of the position in that way. Rather, I will try to paint a nuanced picture of the strengths and weaknesses of the view in the hope that the resulting portrait, when viewed from some distance, will appear attractive and interesting to potential buyers – at least, it will when we bear in mind that any portrait hanging in

the gallery of meta-ethical theories is not without blemishes, and one must balance the blemishes of any portrait against its harmonious aspects.

Aside from illuminating and defending Humean constructivism, I will also try to *improve* the position. I will try to modify the basic theory – as set out by Street (2010: 369–371) – in order to render it a more attractive meta-ethical proposition. (I will elaborate on this modification when I am summarising the contents of the four chapters.) Broadly speaking, the first half of this thesis focuses on traditional, centre-of-the-fairway meta-ethical issues: laying out a particular theory of ethical truth, contrasting it with rival positions, and evaluating its costs and benefits. The latter half of the thesis, by contrast, centres around my proposed modification and will carry us into the realms of moral psychology and phenomenology. So, this thesis involves a mixture of topics – topics that might seem like strange bedfellows – but as I hope to show, there are connecting threads linking them together.

Let me now offer an executive summary of the contents of the four chapters of this thesis. Necessarily, this will be a very compressed summary and will leave out much relevant material; I merely hope to give a sense of the terrain covered.

Chapter 1 focuses on unpacking Humean constructivism, on prodding and probing the position from every angle. I will begin by identifying some recurring motifs among the various strands of constructivism in ethics. I will then draw a general distinction between constructivism understood as a *substantive* normative position – e.g. John Rawls' (1999) theory of justice, which is ultimately compatible with many different meta-ethical theories, be they realist, anti-realist, etc. – and constructivism understood as a meta-normative/meta-ethical one. In §1.1 I will compare and contrast Kantian constructivism and Humean constructivism (both of which are meta-normative/meta-ethical theories).[2] I have already mentioned how the latter position is largely defined *in reaction* to the former, so it is important to get a sense of where the two positions converge and where they diverge.

In §1.2 I will argue against Kantian constructivism. I will offer some reasons for rejecting the position. I should be clear from the outset, however, that it is *not* my

---

[2] I am shading over the distinction between moral reasons and normative reasons; I will explain the reason for this in §1.1. Likewise, in §1.2 I will explain why it is that constructivists focus on *reasons* specifically, as opposed to other facets of ethical truth, such as moral *obligations*.

goal to engage in a long, drawn-out, exhaustive refutation of Kantian constructivism; that would be a lateral move that would distract from the main thrust of this thesis (which is to illuminate, defend, and improve Humean constructivism). Rather, I will try to concisely present some of the key objections to the view. I want to give a sense of why it is that I am putting it to one side and focusing on its more promising meta-ethical cousin.

In §1.3, which is the final section of Chapter 1, I will spend some time further developing Humean constructivism, mapping the contours of the position. I will try to address such questions as: does the view count as a form of moral realism? What exactly are moral judgments under the view, and are they intrinsically motivational? In what sense does the view anchor truths about an agent's reasons to their *actual* conative states? Overall, the content of Chapter 1 is mostly expository and taxonomical in character. This is not to say, however, that it consists of mere recapitulations of what other scholars have written. Rather, it represents my best efforts to provide an original and helpful framework for understanding constructivism in ethics.

In Chapter 2 I will evaluate some of the key strengths of Humean constructivism. I will try to answer the question, why should we be constructivists? What does the position have going for it? Reasons for rejecting *Kantian* constructivism do not automatically equate to reasons for accepting *Humean* constructivism: we need to see some 'positive' reasons in favour of Humean constructivism, in addition to any 'negative' reasons that stem from the fact that its meta-ethical rival is riddled with problems. Chapter 2 is all about exploring these 'positive' reasons. The two main points that I will look at are as follows.

First, despite the fact that Humean constructivism (hereafter: 'constructivism') is an anti-realist, subjectivist view, agents can still be *wrong* about what their moral reasons are under the view. (I will define various technical terms like 'anti-realism' in Chapter 1.) In other words, constructivism allows for a distance between what an agent *thinks* their reasons are and what their reasons *in fact* are. Indeed, this distance can sometimes be considerable, for there are many ways, under a constructivist view of things, in which agents can be mistaken about what reasons they possess. I consider this to be a virtue of the theory, one which I will evaluate in §2.

In §2.1, I will look at the second key point, which is that the theory does a good job of accounting for the practical, action-guiding character of morality; it does a good job of accounting for the *explanandum* that is the 'reliable compresence of moral judgment and corresponding moral motivation' (Kriegel 2012: 471). In doing this, constructivism manages to conserve a key feature of our moral practice. I will contrast the constructivist account of the practicality of morality with that offered by meta-ethical moral realists, such as Brink (1989). I will argue that the realist's notion of moral judgments being ordinary, robust beliefs that track stance-independent moral/normative facts or entities, facts or entities that exist independently of the evaluations of valuing creatures, is problematic. It is problematic because such facts are disconnected from our natural sources of motivation – i.e. conative states – and this makes them vulnerable to the *practical* open question argument, as set out by Blackburn (1998: 70) and Korsgaard (2008: 316).

In Chapter 3 I will evaluate some objections to constructivism. I do not have the space to exhaustively consider every potential objection to the view, so I will focus on a selection of three key objections. In choosing this selection, I have opted for those objections that I perceive to be the most hard-hitting and powerful, objections that I think critics of constructivism would most want to see satisfactorily addressed before they were willing to give the view serious consideration as a meta-ethical theory. I aim to 'pressure-test' constructivism against whatever its keenest critics might throw at it. The first of these objections is one put forth by Russ Shafer-Landau (2003: 41–43), the gist of which is that constructivism tacitly appeals to standards of correctness in the moral domain that are independent of the constructivist procedure, thereby undermining itself. I will unpack this objection in §3.1 and respond to it in §3.2.

My argumentative strategy in dealing with Shafer-Landau's objection is reminiscent of that employed by Sobel (2016: 17–18) in his defence of ethical subjectivism. The strategy is to 'try to diminish the severity of the counter-intuitive … results [that constructivism] generates' – as opposed to denying that it has such results in the first place (ibid.). I will endeavour to show that 'the force of these intuitions [which constructivism purportedly fails to conserve] can be blunted sufficiently to make it sensible to think that the advantages of the view render the overall package acceptable, at least when we keep in mind that any philosophical

account is likely to come with some intuitive costs and we remember that [constructivism's] alternatives are not themselves free from such costs' (ibid.). In short, I will argue that although constructivism is not able to bat aside Shafer-Landau's objection entirely, it is at least able to greatly diminish its force: to defang it, as it were.

The second key objection that I will look at is the amoralist objection. This is adapted from an objection discussed by Sobel (2016: 19–20), but I develop the objection beyond Sobel's construal of it. In greatly abridged form, the objection is that constructivism opens the door to the possibility of *hypothetical amoralists*. An example of a hypothetical amoralist would be a hypothetical agent who is stipulated to be ideally coherent, informed, and imaginative, say, but who, in virtue of his perverse evaluations, has moral reasons that seem utterly perverse and deranged from our perspective (e.g. an agent who has moral reason, by the light of his own perverse evaluative standpoint, to creep up on people in the middle of the night like Freddy Krueger and slash them to pieces).

My discussion of the amoralist objection will encompass §3.3, §3.4, and §3.5. I should note that my argumentative strategy in dealing with this objection is different to that which I use in evaluating Shafer-Landau's objection. Instead of trying to diminish the force of the objection, to defang it, I will argue that constructivism has the resources to neutralise the objection entirely. In other words, the constructivist need not concede that the theory generates counter-intuitive results with respect to this particular objection.

The final objection that I will look at in Chapter 3 is the radical contingency objection. This is based on an objection originally put forth by Dorsey (2018: 579). However, I develop the objection beyond the version articulated by Dorsey in order to make it, in my view, a more potent and hard-hitting challenge to constructivism. I will save the substantive details of this objection for Chapter 3, §3.6. The objection is nuanced and multi-faceted, and it is difficult to do it justice in a bullet-point summary. Suffice it to say that I divide the objection into two halves (this is not something that Dorsey does, but it is, I think, a useful division). I will evaluate the first half of the objection in §3.7 and the second half in §3.8.

In the case of both halves of the objection, my argumentative strategy reverts to that which I used vis-à-vis Shafer-Landau's objection. The strategy is one of frankly

acknowledging that the objection reveals aspects of the theory that are revisionist with respect to at least some of our pre-theoretical intuitions in this domain. However, I will then endeavour to show that these revisionist aspects are not nearly as alarming or eyebrow-raising as they might, at first blush, seem. In other words, I will argue that the revisionist aspects of the theory are *tolerably* rather than *intolerably* so.

Chapter 4, the final chapter, is where I segue from discussing bread-and-butter meta-ethical issues to discussing phenomenology and moral psychology. This is where I begin my experimental attempts to *enhance* constructivism – like the mad scientist in H.G. Wells' (2005) classic novel *The Island of Doctor Moreau*, who attempts to 'enhance' various animal species by producing strange hybrid creatures, like one that is half-man and half-leopard. The enhancement to constructivism that I have in mind is to introduce an element of idealisation into the constructivist methodology for generating truths about one's moral reasons. I want to explore whether we can draw on *peak experiences* in the construction of ethical reality. The rough idea is to restrict the range of evaluations that are fed into the construction to those evaluations that one holds, in either an occurrent or dispositional form, during peak experiences. I call this tentative view 'enlightened constructivism'.

I will unpack enlightened constructivism in far more detail in Chapter 4, §4.3. For now, let me just offer a brief, cursory definition of the term 'peak experiences' so as to make things a bit less abstract. Peak experiences are 'revelations or mystical illuminations that are generally short in duration and usually involve both emotion and cognition'; they are 'moments of highest happiness and fulfilment, and generally carry with them some important meaning and/or insight for the individual' (McDonald 2009: 371). Imagine, for instance, that one climbs to the top of Mount Kilimanjaro and witnesses the morning sunrise from above the clouds. Imagine that at the sight of that sublime sunrise, one is flooded by a feeling of seraphic, otherworldly joy. One feels as if one's mind has grown as bright and expansive as the dawn itself. Now let us say that this feeling of transcendent joy takes on an other-regarding dimension: standing there on the mountaintop like Moses atop Mount Sinai, one starts to feel a boundless, overflowing love towards all conscious creatures, a love that is so intense as to obliterate any sense of egoistic self-concern. Such an experience would be a prime example of a peak experience.

What is of particular interest about such experiences is that when one is in the midst of them, the feeling is not that one's ethical intuitions and values have been deranged; rather, the feeling is one of *heightened* clarity and moral insight. It is as though one has suddenly received a firmware upgrade of one's brain and is able to grasp truths that one had hitherto been cognitively closed to. I think that this aspect of the phenomenology of peak experiences – that they seem, from the first-person point of view of the agent undergoing them, to be profoundly morally and normatively insightful – provides us with an independent reason for trying to incorporate them into our theory of value, whatever that theory turns out to be in the end. This is one of the reasons why I want to try and accommodate peak experiences within the constructivist framework. It is not the only reason, but I will postpone discussing the other reasons until Chapter 4 as they make more sense with the relevant context of Chapters 1 through 3 in place.

This chapter, Chapter 4, is somewhat like a miniature thesis nested within the thesis – it is as though one cracks open an egg only to find a smaller egg contained inside it. This is partly owing to the shift in subject matter from classic meta-ethics to phenomenology and moral psychology. But it is also partly owing to the extraordinary length of the chapter: it is close to half of the length of the entire thesis. The chapter consists of nine sections. These sections are, I think, too mutually entwined for them to be split into independent chapters; or, they could be, but doing so would interrupt their sequential flow.

In §4.1 I define peak experiences. In §4.2 I try to pin down what the felt sense of heightened moral insight during a peak experience could *actually* correspond to under constructivism – e.g. seeing the possibility of a more coherent version of one's 'subjective motivational set', to use Williams' (2008: 60) phrase. In §4.3 I present my 'enhanced' version of constructivism that incorporates peak experiences into the constructivist method for generating truths about one's moral reasons. In §4.4 I evaluate some objections to this view. In §4.5 I explore whether a *further* modification of the theory – i.e. a modification *of* the modification that is enlightened constructivism – might help to address some of these objections.

Without digressing too much on this point, let me just briefly say that what I have in mind regarding this further modification is to tweak the constructivist procedure such that instead of drawing on *conventional* peak experiences in the construction of

truths about one's moral reasons, we draw on *meditative* peak experiences: peak experiences that are induced through meditation practice. The rationale for this is that certain forms of meditation practice can be thought of as systematic methodologies for *reliably* and *consistently* eliciting peak experiences that are ethically salient (in the sense of seeming, from the point of view of the agent undergoing them, to be profoundly morally insightful). This, at least, is the interpretation that I will explore and argue for in §4.6 and §4.7. Just to make this a bit more concrete, an example of an ethically salient form of meditation practice would be *loving-kindness meditation*, which is explicitly geared towards cultivating a state of mind of intense, boundless love, kindness, and compassion, a state that is inherently other-regarding in character (Ricard, Lutz, and Davidson 2014: 43–44). Loving-kindness meditation is one of the key forms of meditation that I will analyse in §4.6 and § 4.7.

In §4.8 I lay out my twice-modified version of constructivism, which I re-baptise 'meditative constructivism'. I consider some independent, theory-neutral reasons in favour of the view. In §4.9, the final section, I evaluate some objections to it. I will conclude that although meditative constructivism represents an interesting and rather quirky position in logical space, the theory faces major unanswered challenges. So, in the end, my schemes for improving the standard, unmodified, non-idealising version of (Humean) constructivism do not quite pan out. This conclusion is not quite as anticlimactic as it might sound. As I hope to show in this thesis, the standard view still has much going for it. And on top of that, the phenomenology of peak experiences – and certain meditative experiences – really does appear to be axiologically relevant in the way that I have described; given that the constructivist tries to build ethical truth using the building blocks that are agents' actual states of valuing, that phenomenology is certainly terrain worth exploring. Indeed, given this peculiar quality that peak experiences have of seeming normatively and ethically insightful, they should be of broader interest to *anyone* who theorises about such matters (even if, in the end, there is no viable way of incorporating them into whatever theory one espouses).

1. Scouting the Terrain: A Bird's-Eye View of Constructivism in Ethics

There is no consensus on how to define constructivism in ethics (Street 2010: 364; Dorsey 2018: 574). But broadly speaking, there are three recurring motifs that we can pick out among the various strands. First, constructivist views try to approach moral theorising in a way that is 'maximally free of any controversial metaphysical suppositions' (Lenman and Shemmer 2012: 2). For instance, constructivists operate with a conception of objectivity in the moral domain that is not metaphysical. That is to say, a conception of objectivity that 'avoids claims to universal and fundamental moral truths that are independent of our fully rational judgments' (Bagnoli 2021: 3).

Second, constructivists put a lot of emphasis on procedures. A constructivist account of some domain of normative facts is 'often [though not always] characterized as one that understands them to be determined by the outcome of some procedure' (Lenman and Shemmer 2012: 2). A classic example of this is the hypothetical procedure in John Rawls' (1999) theory of justice. Rawls sums it up as follows: 'The idea of the original position is to set up a fair *procedure* [emphasis added] so that any principles agreed to will be just ... in order to do this I assume that the parties are situated behind a veil of ignorance ... no one knows his place in society, his class position or social status ... his intelligence and strength ... the parties have no basis for bargaining in the usual sense ... no one is in a position to tailor principles to his advantage' (ibid.: 118–121).

A *non*-constructivist Rawlsian, such as a non-naturalistic moral realist, might say that the original position is a procedure for discovering principles of justice whose truth is *independent* of that procedure. A constructivist Rawlsian, by contrast, would say that there *is* no truth about justice independent of the procedure (Street 2010: 365). Under the latter interpretation, the true principles of justice are not merely uncovered by the original position procedure. Nor do they merely coincide with the outcome of that procedure. Rather, they are constituted by emergence from the original position procedure; emergence from the correct procedure is the truth-maker.

To be clear, I am not claiming that *all* constructivists would sign off on this interpretation. Some constructivists might regard Rawls' original position as a purely

epistemic tool for discovering truths that fall within a particular subset of the larger domain of ethical truth – i.e. truths relating to the principles of justice – as opposed to viewing it as a *constituter* of those truths. For instance, Street (2010: 366) regards the original position as a 'heuristic device whose function is to capture, organize, and help us to investigate what follows from a certain evaluative standpoint on the world – in particular, the evaluative standpoint shared by those of us who accept liberal democratic values such as the freedom and equality of persons'.

The third and final recurring motif is that constructivists usually ground ethical truth in the practical point of view (Lenman and Shemmer 2012: 3). This is the point of view characteristic of a deliberating agent who acts in the world. In the course of deliberating about how to act, agents take certain things/options in the world to be good or bad, right or wrong, better or worse, worthy or worthless, required or optional, valuable or disvaluable, etc. (Street 2010: 366–367). The practical point of view is subject to all of the various motivating states that deliberating agents experience: desires, attitudes of valuing, plans, intentions, normative and evaluative judgments, etc.

Constructivists take certain motivating states – i.e. psychological facts – as raw material and then feed that raw material into a constructivist procedure. What is distinctive about constructivism is that the raw materials in question are *already normative*. What I mean by this is that they are judgments with normative content; I am not suggesting that they possess essentially normative properties. Judgments with normative content can be in an occurrent or dispositional form; they can be implicit as well as explicit. For instance, back in the days of ancient Rome, Cato the Elder asserted that 'to accuse and pursue the wicked was the best thing an honest man could do' (Russell 2004: 226–227). This qualifies as an explicit normative/moral judgment and is exactly the sort of raw material that a constructivist would be interested in. (To be clear, when I use the term 'normative judgment' or 'moral judgment' I am referring to a *mental state* or *mental act*, not a *speech act*.)

The second motif mentioned above – that constructivists usually endorse some hypothetical procedure for determining which principles constitute valid standards of morality – is sometimes taken to be *the* defining feature of constructivism in ethics (Street 2010: 365). Lenman and Shemmer (2012: 3) take it to be so: 'the central,

distinctively constructivist claim is the metaphysical claim about the order of determination moving from procedure to facts and not vice versa'. But Sharon Street, who is arguably the most prominent contemporary defender of constructivism, dissents on this point. According to Street (2010: 365), the proceduralist characterisation fails to capture what is genuinely distinctive about constructivism. The notion of procedure is merely a 'heuristic device'; the philosophical heart of the position is the notion of the 'practical point of view and what does or doesn't follow from within it' (ibid.: 366).

In other words, the constructivist procedure is an epistemic tool for discovering what the entailments of the practical point of view are, but it is the entailments themselves, not our way of finding out about them, that constitute ethical/normative truth.[3] For Street (2010: 366), the *truly* distinctive feature of constructivism is that there are 'no standards of correctness in the normative domain except ... from the point of view of someone who already accepts some normative judgments or other – the point of view of a valuing creature'. In short, Street emphasises the third recurring motif sketched out above, as opposed to the second.

It is Street's practical-point-of-view characterisation of constructivism that I will operate with in this thesis. Before we consider this characterisation in more detail, let us first look at two key reasons that Street gives for rejecting the proceduralist characterisation. The first point is that the proceduralist characterisation seems to confine constructivism to the realm of normative ethics. This is a problem because many constructivists see constructivism as encompassing 'not only a family of substantive normative views but also a unique meta-ethical position' (Street 2010: 365). Imagine that we have two constructivists: Carlos and Charlotte. Carlos claims that valid standards of morality are constituted by agreement among contracting parties characterised as conforming to certain standards of rationality. For Carlos, the constructivist procedure involves multiple individuals and focuses on the negotiations between them. Charlotte, by contrast, claims that valid standards of morality are constituted by emergence from a specific kind of reflective scrutiny (e.g. a particular moral principle is valid if it withstands scrutiny from the full set of a particular agent's normative judgments, together with any relevant non-normative

---

[3] I will elaborate on the notion of 'entailment' and on the distinction between constructivism as a theory of *ethical* truth and constructivism as a theory of *normative* truth further on in this section.

facts). For Charlotte, the constructivist procedure does not involve multiple individuals; it is localised to individual agents.

By the lights of the proceduralist characterisation, it is hard to see how this disagreement between Carlos and Charlotte is anything other than a substantive normative disagreement. It is a disagreement compatible with any number of meta-ethical interpretations, be they realist, expressivist, or whatever else. Carlos and Charlotte are not disagreeing about the meaning of moral statements, about whether moral thinking is, at bottom, continuous or discontinuous with scientific thinking, about what kind of objectivity moral judgments can claim, etc. (Darwall et al. 1992: 140). In other words, their disagreement does not encompass general questions about what we are doing when we engage in moral practice.

The second point here is that the proceduralist characterisation makes it hard to see how constructivism, if it *were* a meta-ethical position, could ever be a *distinct* one. Enoch (2009: 329) sums up this concern: 'the worry is that for any constructivist position, once the details of the constructivist procedure are in place, the constructivist view will be seen to be merely an instance of some other ism (often, but certainly not necessarily, some kind of dispositionalism)'. To flesh this out a bit, Enoch asks us to consider a dispositionalist/response-dependent view of normativity, according to which, 'I have a (normative) reason to do something just in case I would be disposed to have a motivation to do it if I had all (relevant) facts vividly in front of my mind' (ibid.: 328). Under this dispositionalist view, our relevant motivational dispositions do not track an independent order of normative facts. Rather, our dispositions *are* what make it the case that we have certain normative reasons.

Under the proceduralist characterisation, according to which *the* defining feature of constructivism is that constructivists endorse some hypothetical procedure for determining which principles constitute valid standards of morality, the dispositionalist view that Enoch sketches out can easily be construed as a form of constructivism.[4] The process of gathering more information/better quality information and then vividly displaying this information before the mind's eye can be interpreted as a constructivist procedure. The question of what one has reason to do has no answer independent of this procedure.

---

[4] I am focusing on *principles* merely as an example of one key aspect of moral practice.

So, whether or not one has reason to volunteer with one's local conservation group, say, depends entirely on whether being motivated to volunteer with one's local conservation group would be the result of this dispositionalist procedure. The problem here is that this would appear to render the constructivist procedure nothing more than a 'roundabout way of speaking of a straight-forward response-dependence naturalist reduction' or, at best, a 'heuristic device, a way of making a naturalist reduction more intuitively appealing' (Enoch 2009: 328–329). The upshot of the proceduralist characterisation is that *if* constructivism is a meta-ethical position, then it is really just an instance of dispositionalism or naturalism, say.

The key takeaway here is that the proceduralist characterisation is only viable for those constructivists who view constructivism as nothing more than a cluster of substantive normative views or as a form of dispositionalism in disguise. But for those who see constructivism as more than this, as occupying a unique meta-ethical position, it falls short. Those in the former camp are normative constructivists. Those in the latter camp are meta-ethical constructivists. In drawing this distinction, we are distinguishing between substantive normative questions about *which* normative judgments to make and the broader, meta-ethical/meta-normative question of what it *is* to make a moral/normative judgment in the first place. For constructivists of the latter sort, Street's practical-point-of-view characterisation is preferable to the proceduralist characterisation. This is because, as will become clear by the end of this chapter, it allows constructivism to occupy a unique *meta*-ethical position and to weigh in on general questions regarding moral practice, such as questions relating to the *nature* of ethical truth, without committing to a view about its *substantive* content.

As mentioned in the introduction, I am interested in defending a version of meta-ethical constructivism known as 'Humean constructivism', which is a view that originates with Sharon Street (2008; 2010). Humean constructivism is a thesis about the status of ethical truth: it is a thesis about what makes ascriptions of moral reasons true or false (this is a point that I will delve into in more detail in §1.1). What complicates matters somewhat is that the theory doubles up as both a meta-ethical theory and a meta-normative theory. That is, it offers an account of what makes ascriptions of moral reasons true and also an account of what makes ascriptions of normative practical reasons true more generally.

As I will argue in §1.3, it is difficult to separate these accounts as Humean constructivism locates the moral *within* the domain of the normative, so an agent's *moral* reasons are really just a subset of the full set of their normative practical reasons more generally. It is difficult to draw any hard-and-fast distinctions between the moral and the normative under a constructivist view of things because they shade into each other like two colours. I will save further discussion of this point for §1.3. For now, let me just say that I am defining normative practical reasons as reasons to 'pursue, value, or favour something' (Hopster 2017: 769). Such reasons are considerations that, for a particular agent, count in favour of that agent acting in a particular way (Dancy 2006: 45; Finlay 2006: 1; Scanlon 1998: 17). For instance, the fact that a burglar has broken into her house is a reason for Mandy to immediately telephone the police.

In the next section, §1.1, I will compare and contrast Humean constructivism with its close meta-ethical relative: Kantian constructivism. But before I get to that, I want to spend the remainder of this section fleshing out the general distinction between constructivism understood as a substantive normative position and constructivism understood as a meta-ethical/meta-normative position. For the sake of simplicity, I will put the meta-ethical to one side for now and will just focus on contrasting normative constructivism with meta-normative constructivism.

In order to understand this contrast, we must return to Street's notion of the practical point of view, which I referenced earlier. The practical point of view is the point of view occupied by any creature who 'takes at least some things in the world to be good or bad, better or worse, required or optional, worthy or worthless, and so on' (Street 2010: 366). It is the standpoint of a being who judges, either explicitly or implicitly, that some things 'call for, demand, or provide reasons for others' (ibid.). The practical point of view is that of a deliberating agent who acts in the world. In the course of deliberating about how to act, agents make normative judgments. That is, they make judgments with normative content, such as 'Britain would be much better off as a presidential republic than as a constitutional monarchy. If we are going to have a ceremonial head of state, then we ought to have a ceremonial president, just like they do in Germany. That way, the head of state would have some democratic legitimacy, as opposed to being an unelected monarch who is unaccountable to the general public.'

Under *both* normative constructivism and meta-normative constructivism, normative truth consists in what is entailed from within the practical point of view. The term 'entailment' is here meant in a very thin sense. 'Entailment' refers to what follows as a 'purely logical and instrumental matter' from a given set of normative judgments in combination with any non-normative facts that are relevant to those judgments (Street 2010: 367). Instrumental entailment involves the basic principles of instrumental rationality, such as that of means-end coherence. So, for instance, if one judges oneself to have a reason to do something, then one must also judge oneself to have a reason to do what one is fully aware is the necessary means to that end – 'must', on pain of incoherence. Or, to take another example, if one intends to do A and B, then this entails intending to do A. It is important to emphasise that the sense of entailment being appealed to here does not *itself* presuppose any normative notions. To explain this sense of entailment, we need not make any substantive normative assumptions, such as about what anyone '*should* or *ought* to do or infer, or about what counts as a *normative reason* for what' (ibid.).

This is all quite abstract. Let me offer an example to make things more concrete. Florence is planning a trip to the South-West of France. She plans to visit a series of small, rustic villages in the region of Occitanie. These include the villages of Najac, Puycelsi, and Bruniquel, all of which are exceptionally picturesque. Florence is a perfectionist and a very organised person in general; she is planning her trip down to the minutest detail. However, there is one crucial detail that she has overlooked: her passport expired two weeks ago and needs to be renewed if she is to travel abroad. Florence has spent hours researching flights, hotels, restaurants, train tickets, local attractions, etc. In the course of doing all of this meticulous research, she simply took it for granted that something as obvious and essential as her passport did not require any further attention on her part – beyond making sure that it was packed in the suitcase.

We can say that it is *entailed* from Florence's evaluative standpoint that she has a normative practical reason to get her passport renewed before she goes travelling. This follows *given* certain ends that Florence herself endorses as part of her evaluative standpoint (e.g. she wants her trip to France to be as enjoyable and stress-free as possible, she does not want to waste her time or money by being turned away at border control for having an invalid passport, etc.). Importantly, it also

follows *given* certain non-normative facts that are relevant to these ends (e.g. her passport has expired, and one cannot travel abroad with an expired passport).

The difference between normative constructivism and meta-normative constructivism boils down to the following. Normative constructivist views in ethics all begin by specifying some *restricted set* of normative claims (Street 2010: 367–368). They then assert that the truth of a normative claim *falling within that restricted set* consists in that claim's being logically and instrumentally entailed from within the practical point of view, where the practical point of view is given a substantive characterisation. What it means to give the practical point of view a substantive characterisation is that some *further set* of substantive normative claims are assumed, in advance, to be true (this assumption may be tacit rather than explicit).

Basically, all forms of normative constructivism attempt to 'construct some of the judgments constitutive of some proper subset of the normative domain, such as justice or morality, using materials found *elsewhere* [emphasis added] in the practical point of view' (Lenman and Shemmer 2012: 3). The normative constructivist attempts to ground one domain of normative thinking in the materials supplied by another, and by doing so, hopes to provide an account of the truth of some limited subset of the full set of normative claims.

An example of this would be John Rawls' (1999) theory of justice, which I discussed briefly earlier on. Rawls' theory aims to give an account of the truth of those normative claims that concern social and political justice in a liberal democratic society. The truth of normative claims of this particular sort consists in their being entailed from the point of view of the original position. But the crucial thing to note here is that the original position is *itself* value-laden. The hypothetical procedure that is the original position tacitly assumes that certain judgments with normative content are true; it assumes this *in advance of the workings of the procedure*. These are judgments that are 'implicit in the public political culture of a liberal democratic society—including, for example, judgments about the nature of fair bargaining conditions, the freedom and equality of persons, and the irrelevance from a moral point of view of individual traits such as race, sex, class, and natural endowment' (Street 2010: 368). So, Rawls' theory accomplishes its goal, the goal of explaining the truth of some limited subset of the full set of normative claims (i.e. those concerning social and political justice), by appealing to 'what is entailed from

within the standpoint of one who accepts (or whose situation embodies the acceptance of) some *further set* [emphasis added] of substantive normative claims' (Street 2010: 368).

Rawls' view says nothing about what the truth of this further set of substantive normative claims consists in. Because of this, his view is ultimately compatible with any meta-ethical position whatsoever. Metaethical moral realists, for instance, might agree with Rawls that the truth of normative claims concerning social and political justice is constituted by their emergence from the original position procedure. But they might caveat this by saying that the tacit normative assumptions that are baked into the cake of the original position are not *themselves* made true by emergence from any constructivist procedure. Rather, these tacit assumptions are made true by their bearing the correct relation to certain stance-independent moral facts, say.

In contrast to normative versions of constructivism, *meta-normative* constructivism holds that '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' (Street 2010: 369). What it means to give the practical point of view a formal characterisation is that no particular, substantive normative judgments are assumed, *in advance*, to be true. No assumptions – tacit or otherwise – are made about whether certain things/options in the world are good or bad, right or wrong, better or worse, required or optional, worthy or worthless, valuable or disvaluable, etc. To give the practical point of view a formal characterisation is to give an account of the bare standpoint of valuing *as such*; it is to give a *meta* account of what is involved in valuing anything at all. This involves explaining what makes normative judgments true but doing so in a way that does not *presuppose* the truth of any particular, substantive normative judgment (otherwise, it would merely be a form of normative constructivism, like Rawls' (1999) theory).

Meta-normative versions of constructivism do not take the truth of any given substantive normative claim for granted. Such theories offer a general view of normative claims about reasons for action. They explain what a creature must do in order to count as a *valuer* in the first place and how standards of correctness in the normative domain are generated by the attitude of valuing. They are concerned with the *nature* of normative truth, not with its *content*. Under meta-normative

constructivist theories, an agent can, in principle, qualify as a valuer whilst valuing anything at all; one does not have to hold any particular set of substantive evaluations in order to qualify as a valuing creature (e.g. one does not have to tacitly endorse the norms that are implicit in one's society).

At this point, it is worth saying something about terminology. Following Street (2010: 366), I am taking the mental state of normative judgment, of making a judgment with normative content, to be synonymous with the attitude of *valuing*. In both cases, what is being referred to is the conative state of taking certain things/ options in the world to be good or bad, right or wrong, better or worse, worthy or worthless, required or optional, virtuous or sinful, honourable or dishonourable, valuable or disvaluable, etc. It is the state of mind of judging, at least implicitly, that certain things/options 'call for, demand, or provide reasons for others' (ibid.). With this in mind, a valuing creature, according to meta-normative versions of constructivism, is just one who makes at least *some* evaluative judgments: it is a creature who *evaluates* things/options in the world to some degree. These judgments need not be held in an explicit or occurrent form: they could be implicit or dispositional, and they could be about anything.

Throughout this thesis, I will sometimes use the terms 'evaluative attitude', 'evaluation', 'attitudinal valuation', 'state of valuing', 'normative judgment', and 'evaluative judgment' as synonyms for the attitude of valuing. These are all just different ways of referring to the same conative state, as detailed above. I should note that the attitude of valuing, under this construal, is not mutually exclusive with the attitude of desiring. A desire could easily qualify as an evaluation, providing that it has some evaluative component to it, even if only tacit. So, for example, if a citizen of occupied Ukraine has a burning desire to wreak vengeance on the Russian military, then, so long as this desire has an evaluative dimension to it – e.g. 'The way that the Russian military has weaponised torture and bombed civilian targets is a heinous moral evil' – it would qualify as an evaluation. As such, it would be suitable for use as a raw material in the construction of truths about one's reasons (whatever method that involves).[5]

---

[5] In Chapter 3, §3.8 I will explore why it is that the constructivist draws on evaluative attitudes in the construction of ethical and normative reality as opposed to other potentially relevant conative states, such as whimsical wishes or passing fancies.

1.1 HUMEAN CONSTRUCTIVISM VERSUS KANTIAN CONSTRUCTIVISM

Now that we have gotten a sense of the core differences between normative constructivism and meta-normative constructivism, let us turn our attention to the two most prominent versions of meta-normative constructivism: Humean constructivism and Kantian constructivism. As mentioned already, I aim to defend the former position as a thesis about the status of ethical truth. That is to say, a thesis about the conditions under which ascriptions of moral reasons are true. In focusing on moral *reasons* specifically, as opposed to moral obligations, say, I am not tacitly appealing to a form of moral reasons 'fundamentalism' (Schroeder 2021: 3108). I am not implying that all moral facts are grounded in facts about reasons. Rather, I am trying to show that Humean constructivism does a good job of accounting for one key facet of the precious gemstone that is ethical truth: the facet that concerns truths about our moral reasons.

If the theory does a good job of accounting for this facet of ethical truth – and it is yet to be seen whether it does or not – then it might also do a good job of accounting for other facets of the ethical gemstone, such as that which relates to truths about our moral obligations. But this is not an avenue that I will go down in this thesis. I will focus on moral reasons, and my position is one of agnosticism with respect to which kind of ethical truth is fundamental. Truths about moral reasons are more-or-less unanimously considered to be one of the key kinds of ethical truth – and are considered by many to be *the* fundamental kind – so in focusing on them, it is not as though I am making an arbitrary selection (Schroeder 2021).

On the topic of reasons, it is worth briefly addressing another point, which is that I am focusing on *ascriptions of reasons* rather than *reasons themselves*. My reason for doing this is to reduce the need to commit to controversial assumptions regarding the metaphysical nature of reasons – we need not posit any mysterious entities existing 'out there' in the ether, say. This is done in the spirit of the first recurring motif that I mentioned at the very beginning of this chapter, which is that constructivism aims to approach moral theorising in a way that is 'maximally free of any controversial metaphysical suppositions' (Lenman and Shemmer 2012: 2).

With these minor quibbles out of the way, let us proceed to the substance of this section. There are two main forms of meta-normative constructivism: Humean constructivism and Kantian constructivism. The former is a view that originates with

Sharon Street; her statement of the position is the canonical one (Street 2010; 2012). The latter, by contrast, originates with Rawls (1980) but finds its canonical expression in the work of Christine Korsgaard (1996; 2008). I will focus on Street and Korsgaard's construals of these respective positions. As will become clear later on, Humean constructivism is largely defined negatively: it is defined in contrast, in reaction, to Kantian constructivism (Street 2012: 45). As such, the best way of elucidating Humean constructivism is to compare and contrast it with its Kantian cousin.[6]

An analogy we might draw is of two rival sects within Christianity. Humean constructivism is a little bit like the strand of Protestantism that emerged during the Reformation in Tudor England. In order to fully understand it, we need to compare and contrast it with Catholicism in medieval Europe more generally (i.e. Kantian constructivism). To carry this analogy a little further, we can think of Christine Korsgaard as being like the Pope in medieval Christendom (i.e. the highest authority within the Catholic Church). And we can think of Sharon Street as being like the Archbishop of Canterbury (i.e. the highest authority within the Anglican Church in England).[7]

Let me begin by saying what Humean and Kantian constructivists agree on. They agree that the truth of a normative claim consists in its being logically and instrumentally entailed from within the practical point of view, where the practical point of view is given a formal characterisation (i.e. where absolutely no substantive normative claims are assumed, in advance, to be true). They agree that there are no

[6] Obviously, Hume came before Kant, so it may seem a bit odd to frame things this way. The relevant context here is that 'Humean' constructivism is a view that is, at best, only 'vaguely inspired' by Hume, as opposed to being based on a particular reading of his texts (Schroeder 2007: 22). The view is broadly in keeping with a 'time-honored tradition often taken to be vaguely inspired by Hume' (ibid.: 1). *How* exactly it is in keeping with that tradition will become clear by the end of §1.3. For now, let me just say that Humean constructivism is nominally Humean in that it holds that whether a normative practical reason *is* a reason for a particular agent depends, *ultimately*, on some feature of that agent's psychology, such as what that agent actually values. Practical reason *as such* commits us to 'no particular substantive conclusions about our reasons; depending on one's starting set of values, one could in principle have a reason for anything' (Street 2008: 244).

[7] I am leaving aside the complication that it is technically the monarch who is head of the Church of England.

normative truths independent of the deliverances of practical reasoning. Where they disagree is over whether any normative truths are entailed from within the standpoint of valuing *as such*. Kantian constructivists think that we can start with a 'purely formal understanding of the attitude of valuing and demonstrate that recognizably moral values are entailed from within the standpoint of any valuer as such' (Street 2010: 369).

Here is the crux of the matter. For the Kantian constructivist, '*no matter what the particular substantive content* [emphasis added] of a given agent's starting set of normative judgments, moral values are entailed from within that agent's standpoint, whether the agent ever recognizes that entailment or not' (Street 2010: 369–370). Under this view, an agent's moral reasons do not derive from the set of normative judgments that he or she *actually* holds. Rather, they are rooted in his or her nature as a practically rational being. Insofar as moral reasons are justified in terms of the rational requirements that one *must* conform to if one is to qualify as a valuer at all, they universally and necessarily apply to all valuing creatures. So, Kantian constructivism, if successful, vindicates an 'extremely strong form of universalism about reasons, and moral reasons in particular' (Street 2010: 370).

One might wonder what these 'rational requirements' are that the Kantian constructivist sets such great store by. Korsgaard (1996: 121–123) offers the following example. Valuing 'humanity', where 'humanity' is understood as the capacity for rationality, is the condition of the possibility of valuing anything at all. According to Korsgaard, you must be governed by *some* conception of your practical agency, your identity as a *valuing* creature, because if you are not, then 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' (ibid.: 121). But *this* reason for conforming to one's particular practical identity does not *itself* derive from any particular practical identity. Rather, this reason springs from one's identity as a 'reflective animal who needs reasons to act and to live' (ibid.).

Korsgaard's contention is that valuing one's own capacity for rationality, valuing one's own *humanity*, rationally requires that one values the capacity for rationality in others. The capacity for rationality is embodied in all rational beings, so it would be incoherent to value it in oneself but not in others. And valuing the capacity for rationality in others provides one with certain moral reasons – not necessarily

overriding – for acting or not acting in various ways towards those others (e.g. reasons to avoid acting in such a way as to deprive others of practical agency, such as by killing, torturing, or confining them).

For Korsgaard (1996: 121), reasons that derive from our actual, contingently held values are not enough to 'sustain our integrity' when they conflict with the moral reasons that derive from the universal value for the capacity for rationality, say. Bagnoli (2021: 13–14) offers an example of a conflict of this kind:

> [T]he conduct of a Mafioso cannot be coherently justified on the basis of a universal principle [or a universal reason]. The Mafioso thus fails as a rational agent and leads a life that is … not the product of reflective self-government. A systematic failure to be guided by universal principles of self-government amounts to a loss of agency. Insofar as agency is inescapable, we are necessarily bound by the norms of rationality and morality … some kind of integrity is necessary to be an agent and cannot be achieved without a commitment to morality, which is founded on reason.

We can contrast this view of things with that offered by the *Humean* constructivist. The Humean constructivist flatly denies that any universal moral or normative reasons are entailed from within the standpoint of 'pure' practical reason. This is the standpoint of valuing *as such*, where absolutely no substantive practical commitments are assumed, in advance, to be true – tacitly or otherwise. The bare-bones logical and instrumental requirements that govern practical reasoning – which I sketched out in §1 – do not amount to *substantive* commitments under this construal. Rather, they are features that are constitutive of the very attitude of valuing, of making judgments with normative content.

The task of specifying what is constitutive of the attitude of valuing is an exercise in descriptive philosophical analysis rather than substantive normative theorising. In undertaking this task, 'the intuitions we will be calling upon are intuitions about what is and isn't recognizably a case of valuing – *not* intuitions about whether someone who recognizably values something is making a *mistake* in doing so' (Street 2010: 374). If one ignores the logical and instrumental requirements that govern practical reasoning in full consciousness of what one is doing, then one is not making a

mistake about a normative matter; rather, one is simply failing to value (Street 2012: 40).

To make this a bit more concrete, we can draw a parallel between attitudes of valuing and ordinary beliefs. Imagine that someone makes the following claim. 'I sincerely and straightforwardly believe that Homer wrote *The Iliad*, but Homer did not write *The Iliad*.' What are we to make of this? The self-contradictory nature of the statement gives us to think that there is something wrong with the way this person is using the term 'belief'.[8] It is not that this person has a *false* belief that Homer wrote *The Iliad*, but that this person is *failing* to believe that Homer wrote *The Iliad*. This agent is not doing what they need to do in order to qualify as holding a genuine belief of this kind. Examples like this indicate that ordinary beliefs, as a constitutive matter, *aim at truth* (Velleman 2000; Sinclair 2006).

In like fashion, the attitude of judging something to be a reason – which is one form that the attitude of valuing might take – aims, as a constitutive matter, at means-ends coherence. Imagine that someone says, 'I have an overriding reason to read *The Iliad* as soon as possible, and to do so it is necessary that I purchase a copy of *The Iliad*, and I have no reason whatsoever to purchase a copy of *The Iliad*.' It is not that this person is making a *false* judgment about their reasons, but that they are simply *failing* to make a judgment about their reasons. They are not doing what they need to do in order to qualify as making a genuine evaluation.

The Humean constructivist maintains that the substantive content of an agent's reasons is a 'function of his or her particular, contingently given, evaluative starting points [together with the relevant non-normative facts]' (Street 2010: 370). The crucial point is that the bare standpoint of pure practical reason does not in any way determine the substantive content of one's reasons. As Street (2010: 370) puts it, 'the rabbit of substantive reasons … [cannot] be pulled out of a formalist hat [i.e. the standpoint of pure practical reason]: to get substance out, we need to put substance in'.

---

[8] To be exact, the statement does not involve an explicit contradiction: the contradiction is between the implied content of the second conjunct and the explicit content of the first, the implied content being that this person sincerely believes that Homer did not write *The Iliad* (in addition to sincerely believing that Homer *did* write *The Iliad*).

On this view, the value of *humanity*, in Korsgaard's (1996) sense, does not count as a constitutive norm of reasoning. Unlike the norm of means-end coherence (outlined above), the value of humanity is *not* a norm that one must conform to if one is to count as a valuing creature, a creature who takes at least some things/options in the world to be good or bad, right or wrong, better or worse, required or optional, worthy or worthless, valuable or disvaluable, etc. Rather, Korsgaard's value of humanity amounts to a substantive value, a value for the practical agency of other rational beings. It is a substantive value that the Kantian constructivist tries to smuggle into the standpoint of pure practical reason in order to secure a desirable outcome: the universality of certain moral reasons among all rational beings. It is a rabbit that the Kantian constructivist smuggles into the hat so that he can perform the magic trick of pulling it out again. (I will provide some actual arguments to back up these assertions in §1.2. For the moment, I am just trying to give a sense of the key differences between Kantian constructivism and Humean constructivism.)

The Humean constructivist insists that the substantive content of one's reasons must ultimately be supplied by whatever set of evaluative attitudes that one actually, contingently holds (in either an occurrent or dispositional form). If one had different evaluations, one's reasons might be different. And if one undergoes some seismic shift in one's evaluations, then one's reasons might well mirror this change (though whether they do or not depends, at least in part, on what the relevant non-normative facts are). For an example of a seismic shift of this kind, consider the following passage from the Bible, in which Saul undergoes an intense conversion experience on the road to Damascus:

And Saul, yet breathing out threatenings and slaughter against the disciples of the Lord, went unto the high priest, **2** and desired of him letters to Damascus to the synagogues, that if he found any of this way, whether they were men or women, he might bring them bound unto Jerusalem. **3** And as he journeyed, he came near Damascus: and suddenly there shined round about him a light from heaven: **4** and he fell to the earth, and heard a voice saying unto him, Saul, Saul, why persecutest thou me? **5** And he said, Who art thou, Lord? And the Lord said, I am Jesus whom thou persecutest: *it is* hard for thee to kick against the pricks. **6** And he trembling and astonished said, Lord, what wilt

thou have me to do? And the Lord *said* unto him, Arise, and go into the city, and it shall be told thee what thou must do. **7** And the men which journeyed with him stood speechless, hearing a voice, but seeing no man. **8** And Saul arose from the earth; and when his eyes were opened, he saw no man: but they led him by the hand, and brought *him* into Damascus. **9** And he was three days without sight, and neither did eat nor drink … [and once the blind Saul finally arrived in Damascus] immediately there fell from his eyes as it had been scales: and he received sight forthwith, and arose, and was baptized. **19** And when he had received meat, he was strengthened. Then was Saul certain days with the disciples which were at Damascus. **20** And straightway he preached Christ in the synagogues, that he is the Son of God. **21** But all that heard *him* were amazed, and said; Is not this he that destroyed them which called on this name in Jerusalem, and came hither for that intent, that he might bring them bound unto the chief priests? **22** But Saul increased the more in strength, and confounded the Jews which dwelt at Damascus, proving that this is very Christ (The Holy Bible: The Acts. 9: 1–22).

Prior to his conversion experience, Saul had a particular set of evaluative attitudes that involved extreme disapproval of the followers of Christ. So much so, in fact, that his original purpose for going to Damascus was to persecute Christ's followers. But Saul's mystical experience en route to Damascus precipitated a seismic shift in his evaluations. He was bathed in glorious light from on high, and he heard the awesome voice of the Lord, and all kinds of reconfigurations took place in his subjective motivational set.[9]

Under a Humean constructivist view of practical reasoning, the new/altered evaluative attitudes in Saul's motivational set might result in his coming into possession of new/altered normative or moral reasons. For example, it might be that after his conversion experience, Saul no longer has a reason – if he ever did – to

---

[9] This prefigures my discussion of peak experiences in Chapter 4. In that chapter, I will go into far more detail about all of the ways in which peak experiences can alter an agent's evaluations and how that can, in turn, bring about a change in the reasons that an agent possesses. Given that this topic forms the principal subject matter of Chapter 4, I will postpone any further discussion of it until then.

persecute the followers of Christ. Indeed, he might now have a reason to join them, seeing that he has come to believe in their god.

In sum, the key difference between Humean constructivism and Kantian constructivism is that Humean constructivism denies that any moral or normative reasons are entailed from within the standpoint of pure practical reason, absent any substantive evaluations. The bare-bones logical and instrumental requirements that govern practical reasoning – e.g. the principle of means-end coherence – do not amount to substantive values. As such, one cannot derive reasons from them that are universal to all valuing creatures and which apply irrespective of whatever values those creatures *actually, contingently* hold.

Under Humean constructivism, any convergence in the reasons possessed by agents depends, ultimately, on 'contingent similarities in people's evaluative starting points and circumstances' (Street 2010: 370). This might include commonalities in our moral psychology produced by our shared evolutionary history as a species (I will develop this point in considerable detail in Chapter 3, §3.7). The bottom line is that convergence among various practical standpoints is possible, but it is not guaranteed by facts about the nature of reason or the principles of reason that are authoritative for all rational agents (Bagnoli 2021: 18).

1.2 REASONS FOR REJECTING KANTIAN CONSTRUCTIVISM

So, we have gotten a sense of the key differences between Kantian constructivism and Humean constructivism. As mentioned in §1, one of my aims in this thesis is to defend Humean constructivism as a meta-ethical theory. There are many questions that remain to be answered vis-à-vis Humean constructivism. For instance, under the view, does the strength of a reason co-vary with the strength of the evaluations to which it is anchored? Does the view count as a form of moral realism? Does it involve a form of motivational internalism?

Rather than address these questions now, I am going to kick the can down the road and address them in §1.3. The reason for this is that I want to first give some account of why it is that I am choosing *not* to focus on Kantian constructivism in this thesis. As covered in §1.1, Kantian constructivism is a prominent version of meta-normative/meta-ethical constructivism. As such, it may seem strange – or insufficiently motivated – that I am sidelining it and focusing exclusively on its meta-

ethical cousin: Humean constructivism. What I want to do in this section, then, is offer some reasons for rejecting Kantian constructivism. I should be clear from the outset, however, that it is *not* my goal to engage in a long, drawn-out, exhaustive refutation of Kantian constructivism; that would be a lateral move that would distract from the main thrust of this thesis. Rather, I want to try and concisely present some of the key objections to the position; I want to give a sense of why it is not profitable to spend too much time discussing the position.

We can begin by noting that many critics are sceptical about the prospects of Kantian constructivism to derive moral reasons from the bare structure of practical agency, from structural claims concerning rationality (Bagnoli 2013: 9–10). These critics tend to object that 'the Kantian model of rational agency is not rich enough to provide grounds for subjective reasons'; or they 'level against [Kantian] constructivism the objection of formalism, which threatens Kant's ethics'; or they 'doubt that anything substantial can be derived from the sheer logic of rational agency' (ibid.). The points that I raise in this section are mostly in the spirit of the last of these objections: that no reasons with substantive content can be derived from the sheer logic of rational agency.

As mentioned in §1.1, I am focusing on Christine Korsgaard's (1996) seminal construal of Kantian constructivism. Let us consider Korsgaard's core argument for the position (which I referenced in a very cursory way in the preceding section). Korsgaard undertakes to show that if one acts at all in the world, then one must value one's own practical agency, one's own capacity for rationality, and if one values one's own practical agency, then one must – on pain of incoherence – value the practical agency of others, and this entails that one has certain substantive reasons for action concerning those others (ibid.: 123).

Before I enumerate the premises of Korsgaard's argument, it is worth noting that Korsgaard (1996: 123ff.) couches her argument in the language of 'practical identity'. As in, 'any practical identity whatsoever entails identifying with one's humanity … it entails a practical identity as a reflective chooser' (Gibbard 1999: 153). I think that this terminology is a little bit obscure; Korsgaard's argument is clearer and leaves less room for misunderstandings when translated into the language of *valuing*, so I will translate it as such:

(1) **In order to act in the world, one must first take something or other to be valuable.** We need not assume anything about the substantive content of one's values. It does not matter *what* one values so long as one is recognisably valuing something.

(2) **If one does value anything, then one must take that state of valuing to be *itself* supported by further reasons.** This premise is best understood as a claim about what is involved, as a constitutive matter, in the attitude of valuing. The idea is that 'if you value something, then you cannot—simultaneously, in full, conscious awareness—also think that there is *no reason whatsoever* to value it' (Street 2012: 46). Or, to phrase it in more obscure Kantian language, 'a faculty isn't recognizable as the faculty of reason at all if that faculty consciously takes itself to have absolutely no reason whatsoever for what it is doing or thinking at that moment' (ibid.).

(3) **In order for a particular value to be regarded as normative by a valuing agent, there must be some *further* value – also regarded as normative by that agent – that provides a reason for accepting it; but that further value must *itself* have some further value underlying it and which provides a reason for accepting it; and so on, and so on.**

(4) **This regress of supporting values cannot go on forever. If any particular value is to be regarded as normative by an agent, then it must be regarded as being *ultimately* supported by some further value that brings the regress of values to a satisfactory end.** To illustrate this point, imagine that a rational agent, Lord Genji, values the honour of his family and household.[10] If Genji is to regard this value as being normative for him, then he must, when he reflects on the matter, regard it as being ultimately supported by some value that brings the regress of values to a satisfactory end. If he does not think this, if he thinks, on the contrary, that the regress of values bottoms out with *no*

---

[10] See *The Tale of Genji* by Lady Murasaki Shikibu for more detail. *The Tale of Genji* is generally considered to be the supreme masterpiece of Japanese literature (Britannica Academic 2023: 'The Tale of Genji'). It is a huge aesthetic achievement, a profoundly beautiful and evocative tale – and quite saucy as well.

reason, then he will, in effect, regard himself as ultimately having no reason at all for valuing the honour of his family and household. And as a result, this value will cease to be normative for him.

(5) **At the bottom of this regress of values, then, we need a value that: (a) provides a reason for accepting other values; (b) does not *itself* require any reason in order to be accepted.** We can think of this value as a master value: it is the key that unlocks all of one's other values.

(6) **The value for one's own practical agency – i.e. valuing one's own capacity for rationality – when accepted as normative, meets condition (a).** Valuing one's own capacity for rationality provides one with a reason for embracing one's other values as normative. The capacity for rationality is the capacity to distance oneself from one's impulses and desires and to ask, 'Should I act?' Now, if we embrace this capacity as normative, it sets us up with a problem: we need reasons to act (Street 2012: 47). But we can only have reasons if we endorse some particular value or other (at least tacitly). This is because there *are* no reasons independent of the deliverances of practical reasoning (according to meta-normative/meta-ethical constructivism). One can only have reasons to act if one first takes something or other to be valuable: if one makes substantive evaluations. The state of valuing is what sets the standards for determining one's reasons.[11] So, we need reasons to act in the world, and we can only have reasons if we endorse some particular value or other. It follows from this that we have a reason to endorse particular values. To make this a bit more concrete, let us return to the example of Lord Genji. If Lord Genji values his own capacity for rationality, then this provides him with a reason to embrace his particular, contingent value for the honour of his family and household. As an agent endowed with the capacity for rationality, he *needs* some reason or other to act in the world: he cannot act without one. In short, *he has reason to find sources of reasons*. One

---

[11] The background assumption here is the rejection of meta-normative realism. Meta-normative realism is the view that 'there are facts about what reasons agents have for acting, and that those facts are stance-independently true' (Behrends 2013: 486). To say that something is stance-independently true is to say that its truth is 'not dependent on or grounded in the attitudes, either actual or hypothetical, of any agent or possible agent' (ibid.).

source of reasons is his valuing the honour of his family and household. He has a reason to embrace this attitudinal valuation because it will, in turn, provide him with the reasons that he needs in order to act in the world. And so, too, with any other substantive evaluations that he might hold.

(7) **The value for one's own practical agency – i.e. valuing one's own capacity for rationality – when accepted as normative, meets condition (b), as stipulated in premise (5) above (that is, it does not *itself* require any reason in order to be accepted).** If one did not value anything, then one would have no reasons for action. If one had no reasons for action, then one would not be able to act at all. Since one cannot act without reasons, and since one's capacity for rationality is the ultimate source of all of one's reasons, one must value one's own capacity for rationality if one is to act in the world. If one did *not* value one's own capacity for rationality, then none of one's other values would be normative for one. And as a result, one would experience a total paralysis of agency – a truly terrifying prospect. This supplies one with a reason to value one's own capacity for rationality. But, crucially, it is a reason that *springs from that very capacity for rationality, from one's very identity as a reflective chooser who needs reasons and who, therefore, needs some normative conception or other of their identity*. We have hoisted ourselves up by our own bootstraps.

(8) **Valuing one's own capacity for rationality rationally requires that one values the capacity for rationality in others.** The capacity for rationality is embodied in all rational beings, so it would be incoherent to value that capacity in oneself but not in others.

(9) **Valuing the practical agency of other rational beings provides one with reasons to act or refrain from acting in certain ways towards those beings.** For instance, if one values the practical agency of others – which, as established in premises (6) through (8), one *must* do if one makes literally any substantive evaluations at all – then one has reason not to act in such a way as to extinguish, diminish, or constrain the

practical agency of others. Conversely, one has reason *to* act in such a way as to promote and nurture the practical agency of others.

Therefore,

(10) **If one values anything at all, then one must value one's own capacity for rationality, and if one values one's own capacity for rationality, one must value the capacity for rationality in others, and this valuation provides one with reasons for acting or not acting in various ways towards those others; crucially, these reasons apply *universally* and *necessarily* to all practically rational agents.** This is how the Kantian constructivist achieves the 'extremely strong form of universalism about reasons' that I touched on in §1.1 (Street 2010: 370).

This is a long and complex argument. In keeping with the Kantian tradition, it is transcendental in character: it takes something that we cannot act without accepting, derives a consequence from it, and then embraces that consequence. There are many steps in the argument that one might take issue with. I do not have the space to follow up on all of those various leads. Instead, I will focus on two key objections to the argument that I perceive to be the most incisive.

The first objection – which takes aim at premise (7) – is that one can value some particular practical identity *in order* to have reasons to act without also valuing one's identity as a 'reflective chooser' who *requires* reasons in order to act (Gibbard 1999: 153). In other words, one can take certain substantive values to be normative for one in order that one has reasons to act without *also* valuing one's bare capacity for rationality, the very capacity that makes it such that one *needs* reasons in order to act. Consider the following example.

Imagine an Achaean hoplite who is committed to a conception of himself as a courageous, unflappable, skilful, and merciless warrior (Gibbard 1999: 154). 'Achaean', in this context, refers to the ancient Greek warriors in Homer's *Iliad*. Under Korsgaard's (1996) account, acting as befits an Achaean warrior is one way of being a reflective chooser. This is because living up to *any* particular, contingent identity entails being a reflective chooser. (Recall that Korsgaard (ibid.: 123) phrases her core argument in the language of 'practical identity'; all this really involves is

taking at least some substantive values to be normative for one – in this case, values relating to being a valiant warrior.)

So, by Korsgaard's lights, if an agent values himself as a brave Achaean, then he is committed to valuing himself as a reflective chooser. But as Gibbard (1999: 154) points out, why couldn't this Achaean hoplite think of reflective choice as a burden, a burden only mitigated by the laudable way in which warriors like him handle it? The Achaean hoplite might value being someone who deals with his reflective choices valorously and yet still 'disvalue the sheer state of being a reflective chooser' (ibid.).

Gibbard (1999: 154) offers a further example of an ascetic who engages in self-mortification. This ascetic values being someone who 'resists the cravings of the flesh' (ibid.). To add a bit of colour to this example, consider the following extract from *Dubliners* by James Joyce (1992: 201–202), in which several characters are discussing the monks who live in the Abbey of St. Bernard de Trappe in County Waterford, Ireland:

Mrs Malins, who had been silent all through the supper, said that her son was going down to Mount Melleray in a week or so. The table then spoke of Mount Melleray, how bracing the air was down there, how hospitable the monks were and how they never asked for a penny-piece from their guests.

—And do you mean to say, asked Mr Browne incredulously, that a chap can go down there and put up there as if it were a hotel and live on the fat of the land and then come away without paying anything?

—O, most people give some donation to the monastery when they leave, said Mary Jane.

—I wish we had an institution like that in our Church, said Mr Browne candidly.

He was astonished to hear that the monks never spoke, got up at two in the morning and slept in their coffins. He asked what they did it for.

—That's the rule of the order, said Aunt Kate firmly.

—Yes, but why? asked Mr Browne.

Aunt Kate repeated that it was the rule, that was all. Mr Browne still seemed not to understand. Freddy Malins explained to him, as best he could, that the

monks were trying to make up for the sins committed by all the sinners in the outside world.

Okay, so imagine that we have a Christian renunciate who: (a) never speaks; (b) gets up at two in the morning; (c) sleeps in his coffin. Imagine that this person is committed to a conception of himself as a self-flagellating ascetic. He thinks that by living this way he is expunging not only his own sins but the sins of all the saucy hedonists and libertines out there.

As Korsgaard would have it, the fact that this agent is committed to a conception of himself as a self-mortifying ascetic entails that he is committed to his practical identity as a reflective chooser. In other words, it entails that he *positively values* his own capacity for rationality. But this line of reasoning is analogous to saying that because the ascetic values being someone who *resists* cravings of the flesh, and because he cannot be someone who does that without first having cravings of the flesh, he is, therefore, committed to *valuing having cravings of the flesh*. Is the ascetic really committed to valuing this, this very feature that he shares with the heedless pleasure-seekers whom he pities and despises? It seems odd to say so. The same point applies to the example of the valorous Achaean hoplite:

> [T]he warrior on the field just *assumes* that he makes reflective choices, like anyone else, and against this background, he values making them valorously. In this, does he value making reflective choices *as such* [emphasis added], something even the helot or tradesman does? It doesn't seem that his thoughts commit him to this. Whether to put his values in a form that carries this entailment is a question that hasn't arisen for him (Gibbard 1999: 155).

This point carries us nicely to the second key objection to Korsgaard's argument, which takes aim at premises (4) and (5). There is a problem with the very question that Korsgaard sees her argument as answering – and with her conviction that there is a coherent answer to it. The question is, does it really matter whether we act as our nature as reflective choosers requires? Does it really matter whether we accept particular values and then act on the basis of the reasons that spring from those values? Korsgaard's conviction is that we have 'no option but to say yes' (1996:

123). But, as Street (2012: 49) points out, 'as a matter of fact we do have an option here … the proper answer is not to say yes, but rather to reject the question at hand as ill-formulated'.

When translated into the language of normative practical reasons, the question is, 'Is there any reason why I should take something or other to be a reason?' (Street 2012: 49). The problem with this question is that within the meta-normative constructivist framework, there really is no place to stand from which to even pose it. There *are* no standards for determining one's reasons that exist independently of the deliverances of practical reasoning; any such standards are set, ultimately, by one's own judgments about what count as reasons (though, as detailed in §1, these evaluative judgments can take many forms, and their content need not be explicitly about reasons) (Street 2008: 220).

One cannot step outside *all* of one's judgments about what count as reasons – i.e. all of one's substantive evaluations – and ask whether one has some further reason for taking any particular thing to be a reason. As Goldman (2006: 476) puts it, 'just as we cannot get outside our entire system of beliefs to compare them to the world as it is in itself, so we cannot question our entire system of coherent and informed motivations or values, although they can change with new experiences'.

The question of whether one has some *further* reason for taking any particular thing to be a reason is ill-formulated in the way that the question 'Are roses more beautiful?' is ill-formulated. The question does not make sense; it lacks the standard for comparison that would make it sensible to us. A sensible version of the question would be, 'Are roses more beautiful than begonias, cherry blossoms, dahlias, orchids, hydrangeas, camellias, magnolias, peonies, lilacs, plumerias, forget-me-nots, jade vines, or bluebells?'

In asking whether one has a reason to take anything at all to be a reason, one is posing a normative question. And yet, in posing this question, one is stepping back from all of one's substantive evaluations, thereby robbing the question of the very standards that would make it sensible. Either one values something or other, or one does not. If one does, then one has reasons that spring from those values (in concert with the relevant non-normative facts). If one does not, then one has no reasons (Street 2008: 238).

Korsgaard's 'value of humanity' – i.e. the value for the capacity for rationality in all rational beings – is an artifice of language, not a substantive value. It amounts to the claim that 'It is valuable that I take something or other to be valuable' (Street 2012: 50). This claim does not have substantive, non-trivial content, just as 'Are roses more beautiful?' does not have substantive, non-trivial content. This claim – that it is valuable that one takes something or other to be valuable – is asserted without reference to the only standards that *could* possibly determine its correctness or incorrectness. It is not asserted from any particular evaluative standpoint. Rather, it is asserted from *nowhere*. And from *nowhere*, there simply are no normative standards.

A Kantian constructivist might respond to this by saying that these claims are, in fact, asserted from somewhere: the standpoint of 'pure' practical agency, the standpoint of agency itself. What we have done is abstract away from any particular substantive evaluations. But this abstracting process, this process of stripping away *all* contingently held substantive evaluations, 'still leaves us with an evaluative standpoint that can set a substantive normative standard' – or so the Kantian constructivist might argue (Street 2012: 50). The idea, then, is that we have cancelled ourselves down as agents until we occupy the standpoint of agency *as such*. This is the standpoint of a creature who is able to 'distance itself from its unreflective evaluative tendencies, and who needs an answer to the question "What should I do?"' (ibid.). Such a creature, by Korsgaard's (1996: 123) reckoning, *needs* reasons. An agent needs to take something or other to be a reason; a creature with a reflective structure to its consciousness needs some normative conception of its identity or other.

But Korsgaard's use of the term 'need' here is misleading. It trades on a linguistic ambiguity. There are two ways in which we can interpret such claims about needing reasons. On one interpretation, such claims state conceptual, constitutive truths. For example, to be a university graduate, you *need* to have a university degree. On this interpretation, Korsgaard's claim about an agent needing to take something or other to be a reason is merely a statement about what is involved in being an agent. On a second interpretation, however, such claims state *substantive normative truths*. On this interpretation, Korsgaard's claim about an agent needing to take something or

other to be a reason is a statement about that agent having a *normative reason* to take something or other to be a normative reason.

Only the first of these interpretations is plausible. By the lights of meta-normative constructivism, 'an agent is simply defined by the fact that it is a creature that takes something or other to be a reason' (Street 2012: 51). Having a reflective structure to one's consciousness is defined by one's taking something or other to be a reason. To advance the second, normative interpretation of these claims is to say something akin to 'A university graduate has reason to have a university degree.' This second interpretation is implausible. A university graduate who does not have a university degree does not have a *substantive reason* to have a university degree. Rather, they are simply not a university graduate at all.

Likewise, an agent who does not take anything at all – literally anything – to be a reason does not have a substantive reason to take something or other to be a reason. Such an agent is simply failing to be a valuing agent. Street (2012: 51) sums it up: 'The alleged value of "humanity"—i.e. the taking of oneself to have reason to take something or other to be a reason—is not a coherent value, and hence cannot bring the regress of questions to a satisfactory end.' (This is the same regress of questions mentioned in premises (3) through (5) of Korsgaard's core argument.)

The bottom line is that one cannot coherently step back from the entire set of one's interlocking evaluations in one go and then ask, from nowhere, whether this set is correct or incorrect (Blackburn 2010: 17). There simply *are* no independent standards that one can appeal to in order to answer this question (this is the implication of the constructivist rejection of meta-normative realism: more on this in Chapter 2, §2.2). One cannot ask, while simultaneously suspending one's acceptance of any substantive evaluation that would be capable of settling the matter, whether one should endorse one's own set of evaluations, or some other set, or none at all. If one were to do this, one would step outside of the standpoint of agency. One would step into a strange realm where there are no normative facts. One would be an agent without practical agency; one would exist in a kind of suspended animation, like a primordial insect fossilised in amber.

I hasten to add, however, that this view of things does not rule out our asking perfectly sensible questions like 'Do I have a reason to reject the values of Heinrich Himmler in favour of my own?' or 'Were the ethical commitments of the Buddha

superior to my own?' Such questions are fine so long as one is, at least implicitly, posing them from the standpoint of some further set of evaluations – however vague or inchoate – concerning what makes one set of ethical commitments, say, more worthy of endorsement than another.

So, the regress of normative questions that Korsgaard identifies in her argument – premises (3) through (5) – does not end with any substantive value. Rather, it ends with an understanding of the 'exact moment at which normative questions cease to make sense—namely, the moment one divorces oneself from the practical point of view altogether, refusing, for that moment, to take any value for granted' (Street 2012: 52). This is how we cut the Gordian knot that is the endless regress of supporting values. It is futile to seek an Archimedean point from which we can evaluate things from no evaluative perspective whatsoever (Sinclair 2021: 95–96).

Korsgaard's (1996: 123) core argument offers the best chance of deriving from the standpoint of pure practical agency reasons that apply universally and necessarily to all practically rational agents; *any* Kantian constructivist will have to pursue a similar transcendental argumentative strategy if they want to achieve this sort of universalism about reasons (Street 2012: 45). So if Korsgaard's argument does not work – and I hope to have given some good reasons for thinking that it does not – then this carries implications for the plausibility of Kantian constructivism more generally.

One could write an entire thesis just evaluating the objections to Kantian constructivism. My goal in this section was merely to present some of the key objections to the position so as to give a sense of why it is that I am choosing not to focus on it in this thesis. The analogy I would draw is of browsing through books in a bookstore, the bookstore that represents the totality of meta-ethical theories on offer. We have been browsing in the constructivist aisle of the bookstore. We have picked up the book that is Kantian constructivism, had a flick through it, and decided that it is not for us. Instead, we are going to purchase the book that is Humean constructivism and take it home to read.

Obviously, reasons for *rejecting* Kantian constructivism do not automatically equate to reasons for *accepting* Humean constructivism (though the fact that the latter position does not involve the same sorts of transcendental, bootstrapping

arguments as the former is already a point in its favour, I think). We need to see some 'positive' reasons for accepting Humean constructivism, in addition to any 'negative' reasons that stem from the fact that its meta-ethical arch-nemesis, Kantian constructivism, is beset with problems. I will discuss some of these 'positive' reasons in Chapter 2. But before I get to that, I want to spend the remainder of this chapter clarifying a few aspects of Humean constructivism.

1.3 CLARIFYING HUMEAN CONSTRUCTIVISM

Let us begin by revisiting a point that I first broached in §1. Humean constructivism doubles up as both a meta-ethical theory and a meta-normative theory: it offers an account of what makes ascriptions of moral reasons true and also an account of what makes ascriptions of normative practical reasons true more generally. It is tricky to separate these accounts because, under Humean constructivism, the moral domain is nested inside of the broader, normative domain – a little bit like those Russian Matryoshka dolls that nest inside one another. Throughout this chapter, I have somewhat blurred the distinction between moral and normative reasons, sliding fluidly between using both terms. But let us now get a bit more precise about how these two kinds of reason differ.

Under Humean constructivism, X is a normative reason for A to Y iff X is logically and instrumentally entailed from within A's evaluative standpoint at a particular moment in time, the evaluative standpoint being composed of the full set of attitudes of valuing held by A – in either an occurrent or dispositional form – together with any non-normative facts that have a bearing on those attitudinal valuations (recall the example of Florence in §1 and the relevant non-normative fact of her passport being out of date).

The above is a statement of Humean constructivism as a meta-normative theory. However, because the view locates the moral within the domain of the normative, an agent's *moral* reasons are just a subset of the full set of their normative practical reasons more generally. Normative truths – i.e. truths about one's normative reasons – are constructions from the attitude of valuing, and so are ethical truths. There is no fundamental difference between moral reasons and normative reasons. They are not 'normatively independent' and do not belong to 'non-overlapping parts of the normative domain' (Forcehimes and Semrau 2018: 700).

What the distinction between moral and normative reasons amounts to is a general – though far from exceptionless – difference in their substantive content. Moral reasons tend to be other-regarding in character. One's moral reasons generally concern the 'interests or welfare either of society as a whole or at least of persons other than the judge or agent' (Haidt 2003: 853). They concern the well-being and flourishing of conscious creatures *other* than the agent who possesses them. A classic example of a moral reason would be a reason to avoid harming others for one's own amusement or gratification, such as by sadistically torturing them. Pinker (2012: 615) relates the example of a group of rebel soldiers who were encamped on a rooftop in Uganda during a period of civil war in the country and who passed the time by kidnapping local women, tying them up, raping them, and then throwing them to their deaths from the top of the building. In such a case, we intuitively want to say that those agents had very strong moral reasons to avoid acting in the way that they did.

Non-moral normative reasons, by contrast, tend to be self-regarding in character: they tend to be concerned with one's own interests, well-being, flourishing, etc. (Though this is, of course, not to give them any kind of pejorative connotation.) Consider the following example. Imagine that a young lawyer, Alexander, has recently applied for a lucrative new job at a major London law firm. The firm he has applied to is much larger and more prestigious than the firm he is currently working at, and if he were to secure the job, it would effectively double his after-tax income. Alexander has performed well in the application process so far. He has scored highly on various psychometric tests and on the written part of the application. But whether or not he lands the job will depend on how well he performs in the final interview, which is to be conducted by three senior partners at the firm. Now, assuming that Alexander positively evaluates certain things/options in the world, such as progressing in his career, earning a higher salary, working at a more prestigious company, etc., then it seems plausible to say that he has a normative reason to do plenty of preparation for the interview so that he can make a good impression on the interviewers, thereby maximising his chances of securing the coveted position.

It is worth noting that under a constructivist view of things, the strength of any particular moral or normative reason that an agent possesses co-varies with the strength of whatever evaluations in that agent's subjective motivational set go

towards determining the truth about that particular reason. (Hereafter, I will use the term 'constructivism' to refer to *Humean constructivism* specifically, as opposed to any other variant of constructivism that I have looked at in this chapter. Also, whenever I use the term 'normative reason', I am really referring to *non-moral* normative reasons; as already discussed, moral reasons just are normative reasons under a constructivist view of things.)

So, other things being equal, the stronger one's relevant evaluations, the stronger one's reasons. 'Other things' include the relevant non-normative facts, which often have a decisive role to play in determining truths about what reasons one possesses (I will do a deep-dive analysis of this point in §3.8). To return to the example of Alexander, if Alexander really, really wants this job and all the benefits that would accrue from it, if it is his *dream* job, then it seems plausible to say that he has a very strong normative reason – possibly an overriding one – to study hard for the interview so as to maximise his chances of being the successful candidate.

Another point worth noting is that under constructivism, judgments about reasons are only ever true *relative* to a particular agent or to a specified set of agents. Reasons are only ever reasons *for someone*. Take the following judgment as an example. 'Hearing the sound of a gunshot is a reason to run for cover.' This judgment is underdetermined. To know whether it is true, we need to know *whose* reason it is (Hopster 2017: 770). It may be true for Fred, who lives in a war zone, but false for Fran, who lives in a peaceful place.

One interesting nuance regarding the distinction between moral and normative reasons is that ostensibly self-regarding reasons can very easily have hidden other-regarding components to them. This is because our own interests are often entwined with the interests of others. To return to the example of Alexander, let us say that Alexander's parents both died in a car crash a few years ago and that, consequently, he is now the main financial provider in his family. Imagine that he has two much younger siblings, both of whom are his dependants, and that one of these siblings is suffering from bone cancer, which requires expensive treatment. We might now be tempted to say that Alexander has moral reason as well as normative reason to try and secure the new job. His doing so would redound to the benefit of others, others who, let us say, do not have anyone else they can count on.

There are no hard-and-fast distinctions that we can draw, under a constructivist view of things, between moral and normative reasons. There is a general difference concerning their other-regarding content, but this difference is painted in broad brushstrokes and admits of many exceptions and edge cases. As mentioned in §1, my interest is in constructivism as a meta-ethical theory: a thesis about the status of ethical truth: a thesis about what makes ascriptions of moral reasons true. With that in mind, I will focus primarily on reasons that are *other-regarding* in character throughout this thesis, as opposed to reasons that are more self-regarding and self-focused. However, given the porous nature of the boundary between moral and normative reasons, it is inevitable that my remarks will, at times, slip the bounds of the moral and verge on the normative. Suffice it to say that, hereafter, whenever I refer to 'reasons' generically, I am, unless otherwise specified, referring to 'reasons whose substantive content is other-regarding in nature'.

Before I move on to discussing some further points, such as whether constructivism counts as a form of moral realism, let me briefly acknowledge one potential concern with what I have said so far. If an agent's moral reasons are anchored to their actual evaluations, this would seem to imply that if an agent has *no* evaluations of an other-regarding character – i.e. if an agent lacks all moral feeling – then this agent will, consequently, have no moral reasons.

So, to return to Pinker's (2012: 615) example of a group of rebel soldiers in Uganda who whiled away the time by torturing local civilians, if, hypothetically, these soldiers had *literally no* evaluations that could plausibly be construed as moral, then it would seem that they had no moral reasons to avoid committing their heinous crimes (though they might still have had normative reasons, such as if their commanding officer was likely to punish them for their behaviour). This seems like a profoundly counter-intuitive result for a meta-ethical theory to generate. All I will say for now is that this is precisely the objection to constructivism that I will evaluate in §3.3, §3.4, and §3.5, so I will postpone further discussion of it until then. I merely mention this objection to show that I am cognisant of it, cognisant of how a realist about value might bristle at the foregoing remarks.

Two questions remain to be addressed concerning constructivism (at least in this chapter: in the next, I will further expand on the view). Does constructivism count as

a form of moral realism? And does it involve a form of motivational internalism? Let me start with the first question.

There are many, many different ways of defining moral realism (Sayre-McCord 1986: 2–3). I do not want to get mired in a Talmudic parsing of all the definitional complexities in this area. Instead, I will focus on the construal of the realist/anti-realist divide set out by Street (2010: 370); for the purposes of this thesis, it is the most useful and relevant way of framing things: 'The key point at issue between moral realists and antirealists is ... whether things are valuable ultimately because we value them (antirealism), or whether we value things ultimately because they possess a value independent of us (realism) ... in other words, is normativity best understood as conferred or recognised?'.

This carries echoes of Plato's famous Euthyphro dilemma, in which Socrates asks Euthyphro the following perplexing question: 'Is the pious loved by the gods because it is pious, or is it pious because it is loved by the gods?' (*Euthyphro*, 10a). Constructivism falls along the latter (anti-realist) side of this divide – and so, too, do various other influential meta-ethical views, such as Bernard Williams' (2008) account of internal reasons or David Lewis' (1989) dispositional theory of value.

The key point here is that constructivism asserts a 'counterfactual dependence of value on the attitudes of valuing creatures' (Street 2010: 371). So, for instance, if Lorenzo did not value saving the lives of others, he would not have a reason to donate blood. Reason-giving status, value-status, etc., are conferred upon things/options in the world by the attitudinal valuations of valuing creatures, as opposed to being recognised in the world by them. In a world where no one made any evaluations at all, there would be nothing of value (Sobel 2016: 1).[12]

For the constructivist, 'there are no facts about what is valuable apart from facts about a certain point of view on the world [the practical point of view] and what is entailed from within that point of view ... [ethical] truth ... does not outrun what follows from within the evaluative standpoint [i.e. the full set of an agent's substantive evaluations, together with any non-normative facts that have a bearing on those

---

[12] It is worth noting that Street's (2010: 370) construal of anti-realism strongly echoes Sobel's (2016: 1) characterisation of meta-ethical subjectivism: 'Subjectivism maintains that things have value because we value them. Caring about stuff makes stuff matter.' With this in mind, we can understand (Humean) constructivism as not only a form of moral anti-realism but also a form of meta-ethical subjectivism – at least under Sobel's (ibid.) highly influential construal of the position.

evaluations], but rather consists in whatever is [logically and instrumentally] entailed from within it' (ibid.). So, let us understand constructivism as a form of moral *anti-realism*. I will contrast it with a version of moral *realism* in Chapter 2, §2.2.

Onto the second question: does constructivism involve a form of motivational internalism? Motivational internalism is the view that there is a 'necessary connection between moral judgment and motivation' (Svavarsdottir 1999: 162). More specifically, it is the claim that 'necessarily, if an agent makes a moral judgment, then, in normal circumstances, she will possess an appropriate motivation' (Sinclair 2020: 71). Constructivism affirms this claim.[13]

Under constructivism, moral judgments are *not* purely cognitive states like ordinary, robust beliefs (Street 2010: 376). By 'ordinary, robust beliefs', I mean mental states that, as a constitutive matter, aim at truth (Sinclair 2006: 255). For beliefs to aim at truth is for the believer to aim to have the representational content of their beliefs match the state of the world. As Sinclair (2006: 255–256) points out, to 'aim at anything is to be subject to criticism should one fail to achieve that aim'; for beliefs to aim at truth, in the above sense, is for beliefs to be 'subject to criticism just in case their content fails to match the state of the world'. So, given that ordinary beliefs are a particular kind of mental state that aims at truth, they are criticisable when their representational content fails to match up with the state of the world – just as a map is criticisable if it turns out to be inaccurate.

Under constructivism, moral terms are used to express a state of mind that is different from belief (hereafter, whenever I speak of 'beliefs', I am referring to ordinary, robust beliefs in the sense outlined above). For the constructivist, moral judgments are conative states: they are attitudes of valuing: they are consciously or unconsciously held evaluative judgments, judgments about 'what is a reason for what, about what one should or ought to do, about what is good, valuable, or worthwhile, about what is morally right or wrong, and so on' (Street 2006: 110). For example, an eco-activist might judge that we have a moral duty to future generations to transition from using fossil fuels to using fully renewable sources of energy as quickly as possible. And likewise, a disgruntled motorist who is running late for work due to an activist demonstration might judge that eco-activists ought to stop

---

[13] In the next chapter I will analyse this claim in greater detail, contrasting it with a series of subtly different claims.

disrupting the commutes of ordinary people who have nothing to do with the fossil fuel industry.

So, under constructivism, if an agent judges that something is good, right, valuable, worthy, something that they ought to do, something that they have reason to do, etc., then, necessarily, they will be at least somewhat motivated to do or pursue that thing. This is, of course, not to suggest that such a motive would necessarily be sufficient to result in action; the opposing motivational obstacles may simply be too great (Brink 1997: 6). If an alcoholic sincerely judges that he owes it to himself and his family to go to rehab and overcome his addiction, then he will necessarily be at least somewhat motivated to do so. But it may be that his alcohol-induced compulsions are so intense as to rebuff any motivation that comes from this judgment and that he continues on his downward spiral of drinking himself to death.

There is a plausible evolutionary rationale for why it is that attitudes of valuing – a.k.a. moral judgments, under a constructivist view of things – differ from ordinary, robust beliefs in being intrinsically motivational. Broadly speaking, ordinary beliefs evolved to serve the biological function of representing to an organism the way that the world is (e.g. there is a freshwater stream located in the next valley) (Blackburn 1993: 168). Attitudes of valuing, by contrast, evolved to serve the biological function of getting an organism to respond to its circumstances – as represented by its ordinary beliefs – with various kinds of behaviour. Attitudes of valuing evolved to mediate the move from features of a situation to a behavioural reaction; someone with a particular attitude of valuing is 'set to react in some way when an occasion arises, just as someone with a standing belief is set to react to new information cognitively in one way or another' (ibid.).

So, for example, imagine that a mother in a hunter-gatherer society judges that she ought to try and forage some nourishing food for her ailing child, who has come down with a wasting sickness. This mother comes to believe, upon being told by another member of the tribe, that there is a ripe fig tree across the river, close to where the striped gazelles sometimes graze. Her attitude of valuing in this case – that she ought to try and forage some nourishing food for her ailing child – primes her to react in a certain way now that her internal model of the world has been updated. It primes her to wade across the river, say, and to pluck the choicest figs from the tree.

The view that I am gesturing at here sometimes flies by the name of 'Humean psychology' (Kriegel 2012: 471). It is the idea that the human mind broadly divides into two metaphysically independent departments: the cognitive and the conative. Ordinary, robust beliefs fall into the former department; attitudes of valuing fall into the latter. Under Humean psychology, ordinary, robust beliefs are incapable of influencing our actions except in conjunction with conative states, such as attitudes of valuing. Any connection between ordinary beliefs and motivation is contingent rather than necessary. When it comes to motivation, it is fair to say that Humeanism is the dominant view (Rosati 2016: 11–12). Many philosophers – e.g. Sinhababu (2009) – have found the basic Humean picture to be the most plausible. It is the view that I will operate with in this thesis.

The constructivist maintains that moral judgments are attitudes of valuing rather than ordinary, robust beliefs. However, this should not be taken to imply that moral judgments cannot be adequately described as true or false. *Nor* should it be taken to imply that constructivists operate with a 'deflationary' notion of truth; constructivism does not construe moral truth as some sort of elaborate linguistic convention, say (Greenough and Lynch 2006: 2). Rather, it operates with a robust notion of truth and offers a substantive truth condition for moral judgments.

For the constructivist, a particular moral judgment (which can be interpreted as a judgment about what moral reasons one possesses, even if said judgment is couched in the broader language of 'should' or 'ought', say) is true if and only if that judgment's meaning corresponds to its truth condition. Its truth condition will be a fact – or set of facts – about what is logically and instrumentally entailed from within the evaluative standpoint of the agent who makes the judgment; these are moral/normative facts about what the agent in question has reason to do. Recall that an agent's evaluative standpoint is composed of the full set of actual evaluations held by that agent – in either an occurrent or dispositional form – together with any *non*-normative facts that have a bearing on those evaluations.

When I talk about 'correspondence' between the meaning of a particular moral judgment and its truth condition, what I mean is that a moral judgment is true if it judges to be the case what *actually is* the case (i.e. if one judges oneself to have a moral reason, say, that one *actually does* possess – 'actually', in that the fact that

one possesses this reason is entailed from within one's own evaluative standpoint). So, for a given bit of 'extra-linguistic reality' (i.e. a moral/normative fact) to make a given moral judgment true, for it to be a *truth-maker*, the judgment and the fact that makes it true must share the same propositional content: what the fact is a fact *that*, is the same as what the judgment is a judgment *that* (Alston 1996: 38).

Imagine that Manfred renders the following moral judgment. 'I ought to let the town drunk sleep in my barn during the Winter so that he doesn't freeze to death from sleeping in the street in the deathly cold. It wouldn't make much of a difference to me, and it might just save his life.'[14] Under constructivism, the truth condition for this particular moral judgment would be a fact – or set of facts – about what moral reasons for action Manfred possesses with respect to this particular case of the town drunk. The constructivist interprets Manfred's moral judgment as an (implicit) judgment about what he has moral reason to do.[15] This judgment is true if it judges to be the case what *actually is* the case – i.e. if it *is*, in fact, logically and instrumentally entailed from within Manfred's particular evaluative standpoint that he has a moral reason to assist the town drunk in this way. (It might be that, depending on the relevant non-normative facts, there is some far more effective way for Manfred to prevent the man from freezing to death during the Winter, and if so, Manfred's moral reasons for action would, under constructivism, reflect this.)[16]

To be clear, when I talk about moral/normative facts acting as truth-makers under constructivism, I am not suggesting that these facts are intrinsically normative in any kind of controversial metaphysical sense; they are just (natural) facts about what is entailed from one's evaluative standpoint, *given* that one holds certain substantive evaluations (recall from §1 and §1.1 that the constructivist operates with a very thin notion of entailment that *does not make any assumptions, in advance, about the substantive content of one's evaluations*).

---

[14] As discussed in §1, when I use the term 'normative judgment' or 'moral judgment' I am referring to a mental state or mental act, not a speech act.

[15] The constructivist view is that moral judgments need not make *explicit* mention of moral reasons in order for their truth-makers to be (natural) facts about what moral reasons one possesses, *given* the composition of one's evaluative standpoint; moral judgments might be 'couched in the language of value, should, ought, goodness, what makes sense, what's rational, worthwhile, and so on' (Street 2008: 209).

[16] I will discuss the crucial role that the relevant non-normative facts play in determining truths about what reasons one possesses in §2.1 and §3.8.

Constructivism is concerned with knowable moral truths: truths about what moral reasons one possesses, *given* the composition of one's evaluative standpoint. The theory does not posit moral truths that are 'epistemically inaccessible': moral truths that can exist even if there is no possibility of our knowing about them (truths about infinite conjunctions might be an example of truths that are knowledge-transcendent in this way) (Harth 2020: 1093). Moral truths, under constructivism, are by their very nature within our reach (Thompson 2006: 72). This is because they depend, *indirectly and in combination with the relevant non-normative facts*, on the actual evaluations of valuing creatures (which might be held in an occurrent or dispositional form). But just because moral truths are *knowable* under constructivism does not mean that they are not objective. The view allows that moral truths might transcend what we believe now, our everyday thoughts. In other words, agents' moral judgments can still be objectively *wrong*. At the beginning of the next chapter I will explore the various ways in which agents can err in their moral judgments under a constructivist view of things.

<div align="center">***</div>

I hope to have gone some way in this chapter towards illuminating the position in logical space that is (Humean) constructivism and showing how it differs from related positions in this neck of the logical woods, such as substantive normative versions of constructivism and Kantian constructivism. In the next chapter, I will discuss some of the key points in favour of (Humean) constructivism; I will try to show what the position has going for it. In doing so, I will build on many of the points raised in this final section, §1.3, especially relating to moral judgments being intrinsically motivational under the view.

2. Why Be a Constructivist? The Appeal of Constructivism

There are two main plus points of constructivism that I will look at in this chapter. The first is that despite the fact that constructivism is an anti-realist view, it allows that agents can be *mistaken* – sometimes profoundly so – about what their reasons are. In other words, the view allows for the possibility of moral error on the part of an agent. It does not lead to a chaotic, 'anything goes' outcome, where an agent's reasons simply are whatever they *think* they are. My discussion of this point will form the subject matter in §2.1. In this section, I will also discuss the related point that constructivism gives us a pluralistic picture of ethical truth: one can have reasons that are in practical contradiction with one another.

The second plus point I will look at is that constructivism does a good job of accounting for the practical, action-guiding character of morality. It does a good job of accounting for the *explanandum* that is the 'reliable compresence of moral judgment and corresponding moral motivation' (Kriegel 2012: 471). My arguments concerning this point will be partly comparative in nature: it is not *just* that constructivism does a good job of accounting for the practicality of morality, but that it does a much *better* job than realist theories such as Brink's (1989) version of ethical naturalism. My discussion of this point will form the subject matter in §2.2.

2.1 Can Agents Be Wrong About Their Reasons?

Under constructivism, there can sometimes be a considerable distance between what an agent *thinks* their reasons are and what their reasons *in fact* are. What accounts for this? What accounts for it is that an agent's actual mental states – i.e. their subjective judgements concerning what reasons they have – do not *directly* determine truths about what reasons they possess. Rather, constructivism holds that *X* is a reason for *A* to *Y* iff *X* is logically and instrumentally entailed from within *A*'s evaluative standpoint at a particular moment in time, the evaluative standpoint being composed of the *full set* of attitudes of valuing held by *A* – in either an occurrent or dispositional form – together with any non-normative facts that have a bearing on those attitudinal valuations.

This formulation opens the door to two key ways in which agents can be mistaken about their reasons. The first is that agents can be – and often are –

ignorant of the relevant non-normative facts. To see this, consider the following example. Lucy is late for work. Lucy is late for work because she can't find her car keys. She has sifted through the contents of her handbag multiple times but can't seem to find them. It suddenly occurs to her to go and look in her bedroom. She thinks, *I probably left them on the bedside cabinet or on the bureau*. Lucy dashes upstairs and starts looking frantically around her bedroom for the elusive car keys.

We can interpret Lucy as implicitly making the following judgment concerning her reasons. 'I have reason to go and look in my bedroom for my car keys.' But she is mistaken about this. She is mistaken because, unbeknownst to her, the car keys have been underneath the refrigerator in the kitchen this whole time. This is a non-normative fact that is highly relevant to Lucy's predicament, and it is one that she is unaware of. We can plausibly say that what Lucy *actually* has reason to do – by the light of her own evaluative standpoint – is to look underneath the refrigerator for her keys. (In this case, we are talking about a (non-moral) normative reason rather than a moral reason, but the broader point still stands vis-à-vis the crucial role that non-normative facts play in determining truths about what reasons one possesses).

The second key way in which agents can be mistaken about their reasons is that they can – and often do – suffer failures of coherence. Agents can fail to notice how certain evaluations are in tension with one another. They can fail to grasp the full implications of certain evaluations. They can fail to keep in view all of the evaluations that they do, in fact, hold. Hopster (2017: 771) sums it up nicely: 'Human beings are not ideally coherent; we rarely make normative judgments by weighing all the relevant options, reflecting on all the relevant norms, and gathering all the relevant non-normative facts. Instead we apply shortcuts and rules of thumb, we make

decisions under conditions of less than full information, and as a result we regularly make mistakes *by the lights of our own* [emphasis added] evaluative standpoint.'[17]

Consider the following example. Hermann, who takes a keen interest in politics, holds the following set of evaluations. 'A more equal society would be a better society. We have a moral duty to try and redress historical injustices. It is profoundly unfair that some people sleep on feather beds in mansions while others sleep on sheets of cardboard on the cold bare street. It is good and worthy to help those who are less fortunate than oneself. Liberty is one of the highest ends we can strive for, one of the foremost values in the pantheon of human values. We ought to safeguard the rights and freedoms of all citizens. It is wrong to coerce people into doing things against their will.'

It might be that the value of equality is in tension with the value of liberty and that Hermann has never fully appreciated this tension. Let us say that in order to achieve the level of equality that Hermann thinks we ought to strive for, one would have to infringe on individual rights to a great degree. One would, for instance, have to engage in a massive programme of redistribution of wealth, a programme that would include increasing the top marginal rate of taxation from 45 per cent to 75 per cent, forcing wealthy landowners to sell off large parts of their estates, imposing rent control on landlords, blocking high-net-worth individuals from moving liquid assets abroad, etc. It may be that Hermann has never thought through what would be practically involved in actualising his vision of a just society. He has never grasped the trade-offs involved.

So, Hermann might *think* that he has a moral reason, say, to vote for one political party over another in the upcoming general election – e.g. the political party that is campaigning on an agenda of radically increasing social equality and which is

---

[17] Recall from §1 that I am taking the mental state of normative judgment, of making a judgment with normative content, to be synonymous with the attitude of *valuing*. In both cases, what is being referred to is the conative state of taking certain things/ options in the world to be good or bad, right or wrong, better or worse, worthy or worthless, required or optional, virtuous or sinful, honourable or dishonourable, valuable or disvaluable, etc. It is the state of mind of judging, at least implicitly, that certain things/options 'call for, demand, or provide reasons for others' (Street 2010: 366). These judgments need not be held in an explicit or occurrent form; they could be implicit or dispositional. Throughout this thesis I will sometimes use the terms 'evaluative attitude', 'evaluation', 'attitudinal valuation', 'state of valuing', 'normative judgment', and 'evaluative judgment' as synonyms for the attitude of valuing. These are all just different ways of referring to the same conative state, as detailed above.

pledging to enact a series of highly redistributive policies. But depending on the totality of evaluations in Hermann's subjective motivational set, the relative strength of each of those evaluations, and whatever non-normative facts have a bearing on those evaluations, he might actually have a moral reason to do the exact opposite: he might have a reason to vote for the *other* major political party in the upcoming general election, the one that is pledging to safeguard individual rights against an overweening central government.

This is just a vignette. In real life, agents' motivational sets are of staggering, multi-layered complexity. The point I am making is merely that under constructivism, agents are not infallible as to what their reasons are. Rather, they can err in their judgments of what they have reason to do: their judgments can be *wrong* – sometimes profoundly so. In this sense, constructivism is a far cry from 'straightforward subjectivist' theories, where 'if *A thinks* that *X* is a reason for her to *Y*, her thinking makes it so' (Hopster 2017: 770). Constructivism is all the more plausible for allowing for the possibility of moral/normative error in this way. It is important to emphasise, however, that *when* agents err as to their reasons, they err relative to standards of moral and normative correctness set from within their *own* evaluative standpoints. They are failing by their own lights, not according to some stance-independent standard of correctness that floats free of the judgments of valuing creatures.

One interesting complication here is that constructivism involves a form of *pluralism* about reasons (Gill 2014). It allows that truths about what reasons one possesses can be in conflict with each other. One could have incompatible reasons for action, such as a reason to do *X* and a reason *not* to do *X* (e.g. a reason to assist one's elderly relative who is suffering from a terminal and debilitating medical condition to commit suicide and a reason not to do so).

I say only that this *could* be the case, not that it *is* the case with most agents. This pluralism about reasons is derivative of the fact that an agent's evaluations can be in practical contradiction with each other (but not in logical contradiction: I will get to this distinction shortly). But, as already covered, an agent's evaluations are only *partly* determinative of what reasons they possess; the relevant non-normative facts have a crucial role to play as well. So, just because an agent's evaluations are in

conflict with each other does not mean that that agent will have reasons that are in conflict with each other.

To return to the example of Hermann, it could be that Hermann is not, in fact, mistaken in his judgment about what reasons he possesses – at least, he is not wholly mistaken. The conflict in his evaluations – i.e. between deeply valuing equality and deeply valuing liberty – could, depending on what the relevant non-normative facts are, translate into a conflict in his reasons. Hermann could have a reason to vote for one party in the upcoming elections and *also* a reason to vote for another – though this is not to say that such reasons would necessarily be of equal weight.

So, Hermann's judgment that he has a reason to vote for the political party that is campaigning on an agenda of massively increasing social equality might be correct: he does have such a reason. But what Hermann fails to realise, let us say, is that he *also* has a reason – an even stronger reason – to vote for the opposing political party, which is pledging to safeguard individual rights against an overreaching central government. Insofar as Hermann fails to recognise this additional, stronger reason that he has, he is still making a mistake. He is lapsing into moral/normative error. But he is not erring to the same degree as in the previous version of the example, where he imagined himself to have a reason that he did not *at all* possess (i.e. a reason to vote for the pro-equality political party).

I have already mentioned that the pluralistic picture of reasons that constructivism presents us with is rooted in the fact that agents can hold evaluative attitudes – in either an occurrent or dispositional form – that are in *contradiction* with each other. It is worth elaborating on this notion of contradiction. In §1.1 and §1.2 of the previous chapter I went into some detail regarding the logical and instrumental requirements that govern practical reasoning, such as the principle of means-end coherence. I showed that under a constructivist view of things, in order for one to qualify as a *valuer*, a valuing creature, one must conform to these requirements – 'must', on pain of incoherence.

So, for example, valuing an end constitutively involves valuing what one is fully aware is the necessary means to that end, as well as seeking a logically consistent set of ends. Constructivism appeals to the notion of what is logically and instrumentally entailed from within the practical point of view of a particular agent. One aspect of this entailment is the weeding out of any logical contradictions that are

present in the *full set* of an agent's attitudes of valuing (whether those attitudes are held in an occurrent or dispositional form). But this does not involve eliminating any *practical* contradictions that may be present.

An example of a logical contradiction would be the following. An agent judges that eating veal is unethical but also judges that eating veal is not unethical. Let us stipulate that in the case of these two judgments, exactly the same thing/option in the world is being referred to: eating the cooked flesh of a young calf for gustatory pleasure. There is no subtle variation in context between the two judgments that alters the nature of the evaluation. For instance, it is not as though, in the case of the latter judgment, the agent is stranded on a desert island where the only source of protein, for whatever bizarre reason, is veal, and the agent must eat that veal in order to survive until help arrives. The agent is not tacitly judging that eating veal is unethical in ordinary circumstances but not unethical in extraordinary circumstances. Rather, the agent is straightforwardly judging that eating veal is simultaneously unethical and not unethical.

This sort of contradiction, logical contradiction, is a form of incoherence. It is a form of incoherence that is eliminated when it comes to entailment from within an agent's evaluative standpoint. What I mean by this is that any evaluative attitudes in an agent's subjective motivational set that are in logical contradiction with each other do *not* go towards determining truths about what reasons that agent possesses. They are sifted out of the set of that agent's evaluations like weevils being sifted out of a sack of flour.

We can distinguish logical contradictions from practical contradictions (Peach 1953). Practical contradictions occur when two or more of an agent's evaluative attitudes are in tension with each other in the following sense. Acting in such a way as to pursue/realise the object of one state of valuing would conflict with acting in such a way as to pursue/realise the object of some other state of valuing. Imagine, for instance, that an agent holds the following judgments. 'We ought to do what we can to protect the environment and conserve areas of natural beauty. We ought to do what we can to ease the housing crisis – i.e. the severe shortage of affordable housing – in our major cities.'

These two judgments are not in logical contradiction with each other, but they may be in practical contradiction. As in, there may simply be no practicable way to

increase the supply of affordable housing in the country without also damaging the environment to some extent (e.g. by relaxing planning restrictions for new developments on green belt land). The constructivist's notion of entailment does not involve eliminating practical contradictions from an agent's motivational set – if it did, there would hardly be any evaluative attitudes left, for there are so many ways in which said attitudes can come into practical conflict with each other.

2.2 CONSTRUCTIVISM AND THE PRACTICALITY OF MORALITY

I discussed in §1.3 how constructivism is committed to a form of motivational internalism. Motivational internalism is the claim that 'necessarily, if an agent makes a moral judgment [i.e. makes an *evaluation*, under a constructivist view of things], then, in normal circumstances, she will possess an appropriate motivation' (Sinclair 2020: 71). There is a necessary connection between moral judgment and motivation (Svavarsdottir 1999: 162). It is worth taking a moment to drill down on this claim and to distinguish it from a cluster of related claims.

Motivational internalism differs from: (a) moral rationalism; (b) reasons internalism; (c) moral internalism. Moral *rationalism* is the claim that 'a moral obligation to ɸ entails a reason to ɸ' (Björklund et al. 2012: 125). *Reasons* internalism is the claim that 'a reason to ɸ requires, perhaps under ideal circumstances, having a desire to ɸ' (ibid.). *Moral* internalism is the claim that 'a moral obligation to ɸ requires, perhaps under ideal circumstances, motivation to ɸ' (ibid.).

It is important to draw these contrasts because motivational internalists have not always clearly distinguished between having a *motive* to perform some action and having a *reason* to perform some action (Brink 1989: 38). That is, they have not always clearly distinguished their thesis about the connection between moral judgment and moral motivation from a distinct, if related, thesis about the connection between moral judgment and reasons for action. In short, motivational internalism has often been conflated with some form of moral *rationalism*. Constructivism is not committed to moral rationalism: constructivists do not claim that if an agent makes a moral judgment, then, in normal circumstances, he will possess an appropriate reason for action.

60

This is because it is not *individual* attitudes of valuing – a.k.a. moral judgments – that matter under constructivism. Rather, it is the *full set* of one's actual attitudes of valuing that are determinative, in concert with the relevant non-normative facts, of truths about what (moral) reasons one possesses. It is not as though each reason that one has is connected, as if by a piece of string, to individual moral judgments in one's subjective motivational set, such that if one forms a new moral judgment, then one will automatically come into possession of an appropriate reason for action.

Constructivism merely claims that if an agent makes a sincere moral judgment, then they will necessarily have at least *some* appropriate motivation to act in accordance with that judgment. As discussed in § 1.3, this is *not* to claim that such motivation would be of sufficient strength to result in action; it could be overridden by countervailing motives (Brink 1997: 7). Furthermore, the strength of the motivation linked to a moral judgment could vary from agent to agent (Svavarsdottir 1999: 166).

In talking about moral judgments above, I have used the qualifying term 'sincere'. As in, if an agent makes a *sincere* moral judgment, then they will necessarily have at least some appropriate motivation. It is worth dwelling on this notion of sincerity for a moment, for it will play a role in my later arguments in Chapter 3, §3.3. A moral judgment is sincere if it is a genuine, unfeigned evaluative attitude on the part of the agent who holds it, as opposed to one that is held with some degree of dissimulation or pretence. An example of dissimulation or pretence would be if an agent is making a moral judgment in an 'inverted-commas' sense (Hare 1952: 124). That is, making a judgment with *sociological* content rather than *normative* content.

For example, if a psychopathic serial killer says that, 'Spiking people's drinks in bars and then abducting and murdering them is morally wrong', what he may actually mean by this is, 'Spiking people's drinks in bars and then abducting and murdering them is considered to be morally wrong by virtually everyone in the society in which I live – though not by me.' The serial killer's 'moral' judgment here amounts to a descriptive claim rather than a normative one. His judgment can be read in a 'non-evaluative, inverted-commas sense, as meaning that so-and-so falls within a class of actions which is generally held (but not by him) to be obligatory in the evaluative, imperative-entailing sense' (Hare 1952: 165). So, when this serial killer judges that the action he performs is morally wrong, he really only judges that it is an action that everyone else – or virtually everyone else – deems to be morally wrong (Blackburn

1998: 60). This, at least, is one interpretation of what might be going on; I am merely using this example to illustrate the broader point that some supposedly moral judgments are not, in fact, genuine, straightforward, authentic evaluations on the part of the agent. I will return to the issue of inverted-commas judgments in §3.3.

With the preceding clarifications in place, let us now get to the substance of this section. One of the attractions of constructivism is that it does a good job of accounting for the practical, action-guiding character of morality. What I am referring to here is the widely shared, pre-theoretical intuition that there is a strong connection between moral judgments and motivation (Björnsson et al. 2015: 1). This intuition is so deeply ingrained in our moral thinking that we are 'likely to doubt the *sincerity* [emphasis added] of someone who verbally affirms such a judgment yet displays no corresponding motivation' (ibid.). We expect moral judgments to play a practical role in deliberation about what to do (Brink 1986: 25). When one is engaged in such deliberation, the 'confident conclusion that some act would be morally wrong is typically taken to exclude that act from further consideration, whereas the conclusion that it morally ought to be done *typically* [emphasis added] ensues in a decision' (Björnsson et al. 2015: 1). The adverb 'typically' is important here because, as already covered, moral motivation can sometimes be outweighed by competing motivational pressures.

To be precise, the *explanandum* that needs to be accounted for here is not merely that we are on occasion – or even frequently – motivated to do that which we judge to be morally good, right, required of us, something that we ought to do, etc. Rather, what needs to be accounted for is that we are *reliably* motivated to do so – e.g. moral motivation will *reliably shift* to track changes in our moral judgments (Rosati 2016: 12). To illustrate these practical aspects of moral thought and talk, consider the following example.

Imagine that a young woman named 'Briseis' works for Greenpeace. Briseis is a member of the crew aboard a Greenpeace ship that patrols the Southern Ocean to prevent Japanese whalers from engaging in illegal commercial whaling. One day, after weeks of passionately decrying the evils of whaling in conversation with other members of the crew, Briseis does something very strange. As the bell signalling lunch break rings out over the ship's decks, and as the rest of the crew shuffle into

the ship's mess hall for a vegan lunch, Briseis pops up to her cabin to retrieve a culinary delicacy. She comes into the mess hall with some smoked whale meat on a plate and proceeds to sit down and eat it in front of the other crew members. Someone asks her what kind of meat it is, and when Briseis replies that it is Japanese whale meat, everyone in the mess hall falls silent and gapes at her in astonishment.

When confronted with an example like this, we intuitively feel that something has gone very wrong indeed. We feel that if Briseis could sit down and gorge herself on whale meat after having spent weeks enthusiastically engaged in the activity of *preventing* whaling, then she must never have been making sincere moral judgments vis-à-vis whaling. There must have been some dissimulation or pretence on her part. Perhaps she is an undercover journalist who has infiltrated a Greenpeace ship in order to learn more about the organisation and their methods. Or perhaps she is a performance artist carrying out an elaborate social experiment.

Another explanation could be that Briseis is suffering from some kind of neurological impairment that is interfering with the otherwise reliable connection between moral judgment and moral motivation. For instance, Roskies (2003: 55–56) has found that damage to the ventromedial cortex in the brain – which is a part of the brain associated with perception, reasoning, declarative knowledge, emotion, and visceral control – can result in patients having 'particular difficulty in acting in accordance with social mores, despite their retained ability to judge appropriately in such situations'. According to Roskies (ibid.: 57), patients suffering from this particular form of brain damage are generally able to make appropriate moral judgements when queried, and some even retain the 'highest level of abstract moral reasoning', but they do not reliably exhibit corresponding moral motivation.[18]

As already discussed, constructivists are motivational internalists. For constructivists, all moral judgments are intrinsically practical. They are intrinsically

---

[18] I should note that Cholbi (2006: 608ff.) has taken issue with Roskies' interpretation of the data relating to patients suffering from damage to the ventromedial cortex. Cholbi argues that a more plausible construal is that such patients are making moral judgments in an inverted-commas sense (as outlined earlier); they are not making sincere, genuine moral judgments. Cholbi (ibid.: 611) claims that VM patients are comparable to psychopaths in that they 'seem to assimilate moral norms to strictly conventional norms, thus demonstrating a logical competency with a set of norms they do not … [affirm, endorse, etc.] apart from recognising their standing as conventional norms'.

practical because they are intrinsically motivating. They are intrinsically motivating because they are conative states, as opposed to purely cognitive states like ordinary, robust beliefs. Conative states are our natural sources of motivation; they are states that 'have as their function to bring about action that realizes their content, or that dispose one to take such action' (Björnsson et al. 2015: 3).

As covered in §1.3, the constructivist maintains that moral judgments are a certain kind of conative state. They are attitudes of valuing: they are consciously or unconsciously held evaluative judgments, judgments about 'what is a reason for what, about what one should or ought to do, about what is good, valuable, or worthwhile, about what is morally right or wrong, and so on' (Street 2006: 110). The upshot of this is that it is not conceptually possible for an agent to make a *sincere* moral judgment without having at least *some* corresponding motivation – even if only a tiny spurt of motivation that is soon swamped by competing motivations.

To return to the example of Briseis, imagine that Briseis makes the following moral judgment at some point during her time aboard the Greenpeace vessel. 'Whaling is a reprehensible, retrograde, unsustainable practice that simply has no place in the 21st century. It disrupts ecosystems and imposes completely unnecessary suffering on the whales being hunted. As such, we ought to do what we can to eradicate the practice.' Before proceeding, let me briefly reiterate a point made in §1, which is that in using the term 'moral judgment' I am referring to a mental state or mental act, not a speech act. Furthermore, to reduce the need to commit to any controversial assumptions regarding the metaphysical nature of mental state types, we can interpret constructivism as referring to *ascriptions* of moral judgments rather than to moral judgments *themselves*. This is done in the spirit of a point made at the very beginning of this thesis, which is that constructivist views generally aim to approach moral theorising in a way that is 'maximally free of any controversial metaphysical suppositions' (Lenman and Shemmer 2012: 2).

The constructivist explanation of Briseis' case is as follows. If Briseis' moral judgment vis-à-vis whaling is sincere, then it just is an attitude of valuing. If it is an attitude of valuing, then it is a conative state. As a conative state, it will necessarily be accompanied by at least some degree of corresponding moral motivation. And *this* is why we find her behaviour – i.e. coming into the mess hall with a big plate of smoked whale meat and then proceeding to tuck in with evident enjoyment – to be

so mystifying, such an aberration. But if, on the other hand, Briseis' moral judgment is *not* sincere, then it might be something *other* than a conative state, such as an ordinary, robust belief. In this case, we would not expect it to be intrinsically motivating. In other words, we need not be shocked at Briseis' behaviour since her moral judgments regarding whaling were never sincere to begin with.

For instance, it might be that Briseis is actually an undercover right-wing journalist who has infiltrated Greenpeace with the intention of exposing the inner workings of the organisation. In this case, her 'moral' judgment might simply be an ordinary belief with sociological, descriptive content, as opposed to normative content. It is an ordinary belief masquerading as an evaluative attitude. Briseis is mimicking the sincere moral judgments held by her crewmates aboard the Greenpeace ship, who really do judge whaling to be a grave moral evil. She is like a method actor who has given a convincing performance hitherto but who suddenly reveals herself to have been playing a *rôle* this whole time.

In either case, the constructivist is able to explain what is going on in a way that conserves our intuitions regarding the 'reliable compresence of moral judgment and corresponding moral motivation' (Kriegel 2012: 471). It is a virtue of the theory that it can do this. To underscore this point, let us contrast the constructivist account of the practicality of morality with that offered by moral realists.

As discussed in §1.3 of the previous chapter, I am operating with Street's (2010: 370) Euthyphro-style framing of the divide between realist and anti-realists: 'The key point at issue between moral realists and antirealists is ... whether things are valuable ultimately because we value them (antirealism), or whether we value things ultimately because they possess a value independent of us (realism) ... in other words, is normativity best understood as conferred or recognised?'. Constructivism falls into the former camp. For the constructivist, reason-giving status, value-status, etc., are conferred upon things/options in the world by the attitudinal valuations of valuing creatures, as opposed to being recognised in the world by them.

In a nutshell, 'there are no facts about what is valuable apart from facts about a certain point of view on the world [the practical point of view] and what is entailed from within that point of view ... [ethical] truth ... does not outrun what follows from within the evaluative standpoint [i.e. the full set of an agent's substantive evaluations,

together with any non-normative facts that have a bearing on those evaluations], but rather consists in whatever is [logically and instrumentally] entailed from within it' (Street 2010: 370).

Moral realists, by contrast, would want to say that there are 'facts about what [moral] reasons agents have for acting, and that those facts are stance-independently true' (Behrends 2013: 486). To say that something is stance-independently true is to say that its truth is 'not dependent on or grounded in the attitudes, either actual or hypothetical, of any agent or possible agent' (ibid.). So, under moral realism (hereafter: 'realism'), a moral judgment is true if it correctly describes certain stance-independent facts or entities (Korsgaard 1996: 19). Moral judgments represent 'matters of fact about the way the world is' (Bagnoli 2021: 10). They are statements or descriptions of said facts.

Realists, then, are committed to a metaphysical claim: there are stance-independent facts about what is right or wrong, good or bad, what this or that agent has moral reason to do, etc. (Brink 1986: 24). It would be an unprofitable digression to spend too much time taxonomising the various strands of realism; I just want to briefly sketch one distinction that I think is useful. When it comes to the metaphysical claim that there are stance-independent moral facts, one major concern is whether those facts are compatible with a naturalistic worldview. According to naturalism, 'the only facts we should believe in are those countenanced by, or at least compatible with, the results of science' (Sayre-McCord 2017: 7). By the lights of naturalism, if the existence of some putative fact is neither established by nor even compatible with science, then there *is* no such fact. And if a theory requires facts that are incompatible with science, then that alone constitutes a formidable argument against it (ibid.).

To be clear, I am not arguing in favour of naturalism; I merely mention the view because, on this point, realism divides into two main strands: non-naturalistic realism and naturalistic realism. Non-naturalistic realists maintain that moral facts are *not* natural facts and that moral knowledge is *not* simply of a piece with scientific knowledge (Sayre-McCord 2017: 9). But they go on to claim that such facts or entities *do* still exist, despite their incompatibility with naturalism. Naturalistic realists, by contrast, maintain that moral facts are 'either themselves natural facts or are at least appropriately compatible with such facts' (ibid.). Naturalistic flavours of realism

66

can be viewed as attempts to bring a stance-independent conception of normativity into harmony with a contemporary scientific worldview.

It is not my object here to evaluate non-naturalistic realism or naturalistic realism; such a project could easily absorb the length of an entire thesis. I merely aim to show how the constructivist account of the practicality of morality has a leg up over the realist account. But this raises the question of *which* strand of realism to focus on: non-naturalistic or naturalistic. I think it is fair to say that many commentators have found the metaphysical commitments of non-naturalistic realism to be 'ontologically suspicious' when compared with those of naturalistic realism (Olson 2014: 84). Non-naturalistic realists usually hold that moral facts are intrinsically normative in the sense of being intrinsically and objectively reason-giving for all rational agents (Korsgaard 1996: 19, 38).

One charge that is often levelled against the notion of intrinsically normative facts is that such facts would have to be so different from the natural facts with which we are familiar and for which we have good evidence that although we can conceive of their existence, we still have good *a posteriori* reason to doubt it (Olson 2014: 85). The idea here is that intrinsically normative facts would be fundamentally different from anything else in the natural world insofar as they could not be 'exhaustively explained or construed out of whatever the naturalist takes to be the general form of real things'; they would be *sui generis*; they would be ontologically fundamental additions to a naturalistic worldview, a worldview which takes contemporary natural science to be the most accurate guide to the nature of reality (ibid.). To be clear, I am *not* taking a position on whether or not it is plausible to posit the existence of intrinsically normative, stance-independent facts or entities. I am merely registering the observation that non-naturalistic forms of realism are generally thought – rightly or wrongly – to involve controversial metaphysical suppositions. Naturalistic forms of realism, by contrast, are generally thought to avoid such suppositions (Sayre-McCord 2017: 9).

At the very beginning of this thesis I identified three recurring motifs among the various strands of constructivism in ethics. One of these was that constructivist views try to approach moral theorising in a way that is 'maximally free of any controversial metaphysical suppositions' (Lenman and Shemmer 2012: 2). This approach is convergent with that of the naturalistic moral realist. Constructivism and naturalistic

realism both try to steer clear of controversial metaphysics. They both try to ensure that any theoretical suppositions are congruous with a contemporary scientific worldview. With this similarity in mind, it seems appropriate to compare the constructivist – i.e. *Humean* constructivist – account of the practicality of morality with that offered by *naturalistic* realists, as opposed to non-naturalistic realists. Naturalistic (moral) realism and constructivism inhabit the same ballpark of logical space: that concerning naturalistic meta-ethical theories. They are more immediate theoretical rivals. *Non*-naturalistic realism and constructivism are, by contrast, on entirely different continents. It is a case, then, of preferring to make an apples-to-apples comparison rather than an apples-to-oranges one.

There are many different versions of naturalistic realism. I am going to focus on David Brink's (1989) influential and much-discussed version, which he describes as a 'nonreductive form of ethical naturalism' (ibid.: 9). Brink's view is that 'there are moral facts and true moral claims whose existence and nature are independent of our beliefs [and evaluative attitudes] about what is right and wrong … our moral claims not only purport to but often do state facts and refer to real properties … we can and do have at least some true moral beliefs and moral knowledge' (ibid.: 7). What makes this view *naturalistic* is that 'moral facts and properties are constituted by, and so supervene upon (or vary in a lawlike way with), natural and social scientific facts and properties even if moral terms are not definable by natural terms' (ibid.: 9).

Brink's view does not depend on any kind of special meaning relation, such as synonymy, between moral terms and corresponding non-moral expressions. Brink is not claiming that we can deduce moral claims from natural claims or that we can define moral terms in natural terms. He is not claiming that moral facts are *identical* to natural facts. Rather, he is claiming they are *constituted* by natural facts (1989: 9).

The difference between identity and constitution here is modal: if a natural fact exhaustively constitutes a moral fact, then it is consistent that it may not have done, and the moral fact could have been exhaustively constituted by something else (Shafer-Landau 2003: 75–76). For Brink (1989: 9), moral facts are constituted by natural facts of the sort that would typically be cited as evidence for said facts or as reasons for thinking that they obtain. So, for instance, if it is a moral fact that Vladimir

Putin was morally wrong to invade Ukraine, then this fact is constituted by various natural facts relating to the invasion, such as the fact that it has killed or maimed thousands of people and destroyed the livelihoods and property of many more.

For Brink (1989: 7), moral judgments are purely cognitive states: they are ordinary, robust beliefs that purport to state moral facts. Forming a moral judgment involves registering that some moral fact obtains. Brink accepts the Humean theory of motivation, according to which ordinary, robust beliefs are incapable of influencing our actions except in conjunction with conative states, such as attitudes of valuing (ibid.: 8). (As discussed in Chapter 1, §1.3, I am also operating with this view.) For Brink, moral judgments are not intrinsically motivational (and neither are moral *facts*).[19]

Any connection, then, between making a moral judgment and being appropriately motivated is contingent rather than necessary. Another way of putting this is that Brink (1989: 42) is an *externalist* about moral motivation. Externalism is the view that 'moral belief or judgment can exist independently of any motivational force' (Roskies 2003: 52). Externalism denies that there is any 'non-trivial conceptually necessary connection between a person's moral judgment and her motivation to act' (Strandberg 2013: 28).

So, as Brink would have it, there is no necessary connection between an agent sincerely endorsing a particular moral claim, such as 'dumping toxic waste into rivers and lakes is morally wrong', and that agent being motivated – at least to some degree – to act in accordance with that claim. On an externalist picture of moral motivation, moral claims like 'dumping toxic waste into rivers and lakes is morally wrong' really just amount to claims about *what to believe* rather than claims about

---

[19] In order to harmonise the existence of moral facts with a naturalistic worldview, in order to *naturalise* them, Brink has relinquished the notion that moral facts are intrinsically and objectively motivating, reason-giving, normative, prescriptive, etc. In doing so, he has rendered moral facts motivationally inert. Natural moral facts can only count as or entail moral reasons and motivate rational agents if they are supplemented by extrinsic sources of motivation, such as a generic, *de dicto* desire on the part of an agent to do whatever is morally right and to refrain from doing whatever is morally wrong. My concern in this chapter is with the link between moral *judgments* and motivation, not with the link between moral *facts* and motivation. Having said this, under Brink's (1989: 7) view, a moral judgment is just an ordinary, robust belief that some natural moral fact obtains; so, in discussing the link between moral judgments and motivation, I cannot avoid indirectly discussing the natural moral facts that moral judgments purport to state.

what to do (Silverstein 1994: 124). The externalist represents moral propositions as propositions to be *believed*.

This is the case even when moral propositions take the form of propositions about what one morally ought to do. Propositions about what one morally ought to do are not actually claims about what to *do* but rather about what to *believe* (Silverstein 1994: 124). Under such a view, questions about what to do and why to do it become external to moral theorising. To judge that a particular act or intention is morally right or wrong, say, is not to make a claim in favour or in opposition to it but merely to make the factual claim that that particular act or intention either does or does not possess a certain property (ibid.: 124–125). In short, Brink's moral theory is a theory about what to believe, not about what to do.

It is not my goal here to comprehensively evaluate Brink's moral theory. My aim is merely to emphasise that constructivism does a good job of accounting for the practical, action-guiding character of morality by contrasting the constructivist account with that offered by Brink (I am taking Brink's theory to be an emblematic form of naturalistic moral realism). I think the most incisive way of presenting the concern about the *impracticality* of moral judgments under Brink's view is through the *practical* open question argument, as set out by Blackburn (1998: 70) and Korsgaard (2008: 316).

If a moral judgment is just an ordinary belief about the way that a particular part of the natural world *is* – i.e. the part concerning natural moral facts – then, even if the belief were 'settled', we would still face the open question of 'what importance to give it, what to do, and all the rest' (Blackburn 1998: 70). Merely registering that a particular natural moral fact obtains still leaves us with the all-important question of 'what to do about it' (ibid.). Even if we *know* that a particular action is morally good, say – i.e. if our beliefs about an action accurately track certain stance-independent moral facts – so long as those beliefs are just 'piece[s] of knowledge', that knowledge still has to be 'applied in action by way of another sort of norm of action, something like an obligation to do those actions which we know to be good' (Korsgaard 2008: 316). And there is no way to 'derive such an obligation from a piece of knowledge that a certain action is good', say (ibid.). In other words, we

cannot derive practical norms of action from purely factual knowledge about moral reality.

This could lead to a situation where an agent becomes alienated from their own moral judgements (i.e. ordinary, robust beliefs, under Brink's view). They might possess genuine moral knowledge regarding stance-independent natural moral facts, but because these pieces of knowledge are severed from that agent's natural sources of motivation – i.e. their conative states – they might be led to ask, 'So what? What is that knowledge to me?'

There is a danger, under Brink's realist view, that pieces of moral knowledge become mere academic *curios* with no practical import, like rare stamps that a stamp collector might save up in a book and look at every now and then, only to then lock the book away in a drawer of his desk when he goes off to do other things (i.e. to *act* in the world). This is not just a psychological question about the impact that cognisance of a natural moral fact would have on one's mind; it is a normative question about the *authority* of such knowledge under a realist view like Brink's (1989).

For the constructivist, by contrast, the practical open question argument does not pose a problem. The constructivist can close shut the open question. This is because, under constructivism, moral judgments are not purely cognitive states like ordinary, robust beliefs. Rather, they are conative states: they are *evaluative* states, states that, at least tacitly, involve judgments about 'what is a [moral] reason for what, about what one should or ought to do, about what is good, valuable, or worthwhile, about what is morally right or wrong, and so on' (Street 2006: 110).

Such judgments have normative meaning and motivational power for the agent who sincerely makes them. Moral facts, under a constructivist view of things, are facts about what moral reasons one possesses. But these facts are not stance-independent. Rather, they very much depend on the composition of one's evaluative standpoint, one's *stance*. Moral facts are constructed, at least in part, out of the raw materials that are one's *actual* moral judgments. Moral knowledge – i.e. knowledge about what moral reasons one possesses – is not dislocated from one's natural sources of motivation (one's conative states) but is built using those natural sources.

It does not make sense to ask, 'So what? Who cares, man? What is that knowledge to me?' when confronted with moral knowledge under a constructivist

view of things, because that knowledge is built out of judgments about what one *does* actually care about, about what *does* actually matter to one. Let me offer an example to illustrate this point.

Imagine that a primary school teacher, Charmian, is very fond of children and is passionate about helping children to achieve their full potential, especially children who have grown up in poverty or in dysfunctional homes and who lack many of the advantages in life that others enjoy. Although Charmian does not earn much from her teaching work, she sets aside some money each month to donate to a handful of children's charities. Among these charities is one by the name of 'Organisation for Ending the Abuse of Children' – 'OEAC' for short. OEAC is focused specifically on rescuing children under the age of ten who are the victims of physical or sexual abuse. We can imagine Charmian making a moral judgment like, 'The abuse of young children is one of the vilest crimes in our society, and we all have a moral duty to prevent it.' As an auxiliary judgment to this, Charmian judges that she has moral reason to donate money to OEAC, which she believes to be one of the most effective charities operating in this area.

Unbeknownst to Charmian, OEAC is actually a corrupt and toxic organisation. The CEO of the charity, Stefan, is a psychopathic paedophile who pockets much of the money that comes in through donations and who spends the rest of it on marketing and social media campaigns so as to make the organisation *appear* like a respectable and well-run charity. In many ways, it is the perfect cover story for Stefan: he is a psychopathic paedophile who enjoys preying on vulnerable children, and where better to do that than as the CEO of a charity that is ostensibly aimed at *preventing* the abuse of children? So, Charmian *thinks* that by donating money to OEAC she is doing something positive to prevent the physical and sexual abuse of minors. But, in reality, her money is being used by a highly malevolent agent to *further* the physical and sexual abuse of minors.

In the case of Charmian, we can plausibly say that she has a moral reason to immediately stop donating money to OEAC – and, presumably, to report the organisation to the police. Now, this truth about what she has reason to do is a *fact*: it is a piece of moral knowledge. But, unlike in the case of the realist, it is *not* a piece of moral knowledge that is disconnected from her natural sources of motivation. Rather,

it is a construct from her conative states (together with the relevant non-normative facts).

When it comes to moral knowledge under constructivism, there is no risk of an agent facing a practical open question when confronted with pieces of that knowledge. This is because moral knowledge is constructed using the raw materials found in one's *own* evaluative standpoint (i.e. the full set of one's attitudes of valuing – held in either an occurrent or dispositional form – together with any non-normative facts that have a bearing on those attitudinal valuations). For Charmian, the fact that she has a moral reason to immediately stop donating money to OEAC is a fact that is intimately bound up with what she cares most about in the moral domain (e.g. preventing the abuse of children, helping children to grow, flourish, and achieve their potential, etc.). And the same holds for any other constructed moral facts about what she has reason to do.

<p style="text-align:center">***</p>

In this chapter, I hope to have given a sense of some of the key attractions of constructivism. This is a relatively short chapter, the shortest in the thesis. This should not be taken to mean that constructivism is lacking in attractions. Rather, the brevity of this chapter is due to the fact that I perceive the project of *parrying objections to constructivism* – which forms the subject matter of the next chapter – to be of greater importance than the project of extolling the virtues of the position. As Sobel (2016: 8) notes, the reason that subjectivistic meta-ethical theories like constructivism are often given short shrift in the literature is not because they lack pros, attractions, upsides, etc. Rather, it is because such views – views which anchor truths about one's moral reasons, say, to the actual conative states of individual agents – are thought to generate hopelessly counter-intuitive results.

In the next chapter, then, I will endeavour to show that the ostensibly counter-intuitive results that constructivism generates are either: (a) illusory (i.e. not, in fact, counter-intuitive) or (b) tolerably rather than intolerably revisionist. In so doing, I hope to 'soften up' the opposition to constructivism (as opposed to spending more time talking up the benefits of the view).

3. AGAINST CONSTRUCTIVISM: THREE KEY OBJECTIONS TO THE POSITION

In this chapter, I will evaluate a selection of three key objections to constructivism. In choosing this selection, I have opted for those objections that I perceive to be the most hard-hitting and powerful, objections that I think critics of constructivism would most want to see satisfactorily addressed before they were willing to give the view serious consideration as a meta-ethical theory. I aim to 'pressure-test' constructivism against whatever its keenest critics might throw at it – will the boiler that is constructivism survive the pressure test, or will it explode?

I will begin with an objection put forth by Russ Shafer-Landau (2003: 41–43). I will unpack the objection in §3.1 and respond to it in §3.2. Shafer-Landau's objection is not actually focused on *Humean* constructivism specifically; it is targeted at meta-ethical constructivist views more broadly. However, with some minor tweaking, the objection can be made to apply to Humean constructivism (hereafter: 'constructivism'). The objection is an elegant and powerful one, and for that reason, it is worth considering.

My argumentative strategy in dealing with Shafer-Landau's objection is reminiscent of that employed by Sobel (2016: 17–18) in his defence of ethical subjectivism. The strategy is to 'try to diminish the severity of the counter-intuitive … results [that constructivism] generates' – as opposed to denying that it has such results in the first place (ibid.). I will endeavour to show that 'the force of these intuitions [which constructivism purportedly fails to conserve] can be blunted sufficiently to make it sensible to think that the advantages of the view render the overall package acceptable, at least when we keep in mind that any philosophical account is likely to come with some intuitive costs and we remember that [constructivism's] alternatives are not themselves free from such costs' (ibid.). In short, I will argue that the constructivist has the resources to if not actually deflect Shafer-Landau's objection, then at least to greatly diminish its force, such that it strikes a mere glancing blow.

The second key objection that I will look at is the amoralist objection. This is adapted from an objection discussed by Sobel (2016: 19–20), but I develop the objection beyond Sobel's presentation of it. I will argue in §3.3 that the amoralist objection really consists of two distinct, separable objections; I will evaluate the first

of these in §3.4 and the second in §3.5. My argumentative strategy in dealing with the amoralist objection – both versions of it – is different to that which I use in evaluating Shafer-Landau's (2003) objection. Instead of trying to diminish the force of the objection, to defang it, I will argue that constructivism has the resources to neutralise the objection entirely. In other words, the constructivist need not concede that the theory generates any counter-intuitive results with respect to this particular objection.

The final objection that I will look at is the radical contingency objection. This is based on an objection originally put forth by Dorsey (2018: 579). However, I develop the objection beyond the version articulated by Dorsey in order to make it, in my view, a more potent and hard-hitting challenge to constructivism. I will argue in §3.6 that the radical contingency objection has two halves to it and that Dorsey only emphasises one half. I will evaluate the first half of the objection in §3.7 and the second half in §3.8.

In the case of both halves of the objection, my argumentative strategy flips back to that which I use in discussing Shafer-Landau's objection. The strategy is one of frankly acknowledging that the objection reveals aspects of the theory that are revisionist with respect to at least some of our pre-theoretical intuitions in this domain. However, I will then endeavour to show that these revisionist aspects are not nearly as alarming or eyebrow-raising as they might, at first blush, seem. In other words, I will argue that the revisionist aspects of the theory are *tolerably* rather than *intolerably* so.

Before I proceed to the substance of this chapter, let me just briefly echo a point made in the introduction to this thesis, which is that it is not my goal in this thesis to try and isolate a single consideration that is so powerful that it secures the truth of constructivism regardless of its other costs. There *is* no single argument that conclusively establishes the truth of the position in that way. Rather, I am trying to paint a nuanced picture of the strengths and weaknesses of the view in the hope that the resulting portrait, when viewed from some distance, will appear attractive and interesting to potential buyers – at least, it will when we bear in mind that any portrait hanging in the gallery of meta-ethical theories is not without blemishes, and one must balance the blemishes of any portrait against its harmonious aspects.

3.1 Unpacking Shafer-Landau's Objection to Constructivism

Russ Shafer-Landau (2003: 41) expresses the following concern about constructivist procedures: '[I]t may be impossible to craft a set of constraints on attitude formation such that the emerging attitudes yield prescriptions that match up with our views about what constitutes paradigmatically moral and immoral behaviour'. This concern is mainly directed at what Shafer-Landau (ibid.: 39) refers to as 'objectivist' versions of constructivism. These are versions of constructivism that require some degree of idealisation with respect to the attitudes and responses that go towards fixing ethical truth. Examples include Firth's (1952) ideal observer theory, Korsgaard's (1996) Kantianism, and Scanlon's (1998) contractualism. Under these 'objectivist' accounts, ethical truth is constructed, for instance, from the 'views that agents take from a perfectly informed and dispassionate standpoint, from the standpoint of pure practical reason … from the perspective of one looking to identify principles that would govern persons on a basis that no one could reasonably reject [etc.]' (Shafer-Landau 2003: 39–40).

Shafer-Landau's concern is that no matter which way we configure the idealisation, the ethical truths generated by the constructivist procedure are unlikely to conserve some of our core intuitions regarding paradigmatic cases of moral and immoral behaviour. For example, agents that satisfy all of the idealising criteria might still be perfectly callous – a bit like Spock from *Star Trek*. And the principles they endorse might seem harsh, clinical, or impossibly demanding from the perspective of us mere mortals. Essentially, this is a concern about extensional adequacy. It is generally considered to be a virtue for a theory to be able to capture our strongest pre-theoretical intuitions about certain cases (Behrends 2016: 13). If constructivism fails to 'render the correct verdicts about what [moral] reasons there are in important cases', then it fails to secure a theoretical virtue that is desirable for any meta-ethical theory to possess: extensional adequacy (ibid.: 1).

This thesis is focused on *Humean* constructivism specifically, so I will adapt Shafer-Landau's objection to that particular theory. Under Shafer-Landau's (2003: 39) taxonomy, Humean constructivism is *not* an 'objectivist' moral theory. Rather, it is a 'subjectivist' one because it takes at least some of the actual and uncorrected attitudes of agents to be determinative, at least in part, of ethical truth (ibid.). (Recall from §1.1 that I am focusing on one key facet of the precious gemstone that is

ethical truth: the facet that concerns truths about our moral *reasons*. I do not discuss other facets of this gemstone, such as that which relates to truths about our moral obligations.)

One minor complication here is that Shafer-Landau's (2003: 42) objection assumes that constructivism involves a *hypothetical procedure* for generating ethical truths. I argued in §1 that for *meta*-ethical versions of constructivism, such as Humean constructivism, Street's (2010) practical-point-of-view characterisation of constructivism is preferable to a proceduralist characterisation. Just as a reminder, the proceduralist characterisation of constructivism has it that 'the central, distinctively constructivist claim is the metaphysical claim about the order of determination moving from procedure to facts and not vice versa' (Lenman and Shemmer 2012: 3). A classic example of this is the hypothetical procedure in John Rawls' (1999: 118) theory of justice: 'The idea of the original position is to set up a fair *procedure* [emphasis added] so that any principles agreed to will be just.'

When it comes to meta-ethical versions of constructivism, like Humean constructivism, the proceduralist characterisation fails to capture what is genuinely distinctive about the view (Street 2010: 365). The notion of procedure is merely a 'heuristic device'; the philosophical heart of the position is the notion of the 'practical point of view and what does or doesn't follow from within it' (ibid.: 366). What is *truly* distinctive about meta-ethical constructivism is that there are 'no standards of correctness in the normative domain except ... from the point of view of someone who already accepts some normative judgments or other – the point of view of a valuing creature' (ibid.). In other words, the constructivist procedure is a purely epistemic tool for discovering what the entailments of the practical point of view are, but it is the entailments themselves, not our way of finding out about them, that constitute ethical/normative truth.

Humean constructivism does not make reference to any kind of procedure – even just as an epistemic tool or heuristic device. Rather, the central notion is of *entailment* from the practical point of view. The view holds that *X* is a reason for *A* to *Y* iff *X* is logically and instrumentally entailed from within *A*'s evaluative standpoint at a particular moment in time, the evaluative standpoint being composed of the *full set* of attitudes of valuing held by *A* – in either an occurrent or dispositional form –

together with any non-normative facts that have a bearing on those attitudinal valuations.

Recall that an attitude of valuing is the conative state of taking certain things/ options in the world to be good or bad, right or wrong, better or worse, required or optional, worthy or worthless, valuable or disvaluable, etc. It is the state of mind of judging, at least implicitly, that certain things/options 'call for, demand, or provide reasons for others' (Street 2010: 366). These judgments need not be held in an explicit or occurrent form; they could be implicit or dispositional. As discussed in §1 and §1.1, the notion of *entailment* that the Humean constructivist operates with is understood in a very thin sense: it appeals to the bare-bones requirements that govern practical reasoning, such as the principle of means-end coherence; these are requirements that any agent must conform to if they are to qualify as a valuing creature at all.

When it comes to Shafer-Landau's (2003) objection to constructivism, nothing of real substance hinges on the distinction between the proceduralist characterisation and the practical-point-of-view characterisation; I mention it merely for the sake of clarity. Shafer-Landau's objection revolves around the point that if we allow the attitudes of agents to go towards fixing ethical truth – even if only indirectly – those attitudes are unlikely to 'yield prescriptions that match up with our views about what constitutes paradigmatically moral and immoral behaviour' (ibid.: 41). Under *either* a proceduralist or practical-point-of-view characterisation of constructivism, the attitudes of agents to go towards fixing ethical truth, so Shafer-Landau's objection is relevant in both cases.

As already mentioned, Shafer-Landau appears to operate with a hypothetical proceduralist characterisation of constructivism (2003: 42ff.). In adapting Shafer-Landau's objection to the case of Humean constructivism specifically, I will modify his terminology. Instead of talking about a 'procedure' for generating truths about one's moral reasons, I will talk about 'logical and instrumental entailment from within the practical point of view'. I fear that it would become quite verbose and mind-numbing to keep using this long phrase over and over again, so what I will do is use the term 'method' as a euphemism for 'logical and instrumental entailment'. As in, I will refer to the Humean constructivist 'method' for generating truths about one's moral reasons.

This 'method' consists of nothing more than logical and instrumental entailment from a given agent's evaluative standpoint (i.e. the full set of attitudes of valuing held by that agent – in either an occurrent or dispositional form – together with any non-normative facts that have a bearing on those attitudinal valuations). Not to belabour the point, but, once again, it is the entailments from within the practical point of view that constitute ethical/normative truth under Humean constructivism, as opposed to emergence from a correct hypothetical procedure; emergence from a correct procedure is not the truth-maker in this particular variant of constructivism.

With these clarifications in place, let us return to unpacking Shafer-Landau's objection. As already mentioned, the objection is directed at what Shafer-Landau (2003: 39) refers to as 'objectivist' versions of constructivism, which are versions of constructivism that require some degree of idealisation with respect to the attitudes and responses that go towards fixing ethical truth. Humean constructivism does not qualify as an 'objectivist' moral theory under this taxonomy. Rather, it is a 'subjectivist' one because it takes at least some of the actual and uncorrected attitudes of agents to be determinative, at least in part, of ethical truth (ibid.). Under Humean constructivism, it is the *actual* attitudes of *actual* agents that go towards fixing ethical truth, as opposed to the hypothetical attitudes of hypothetical agents.[20]

Shafer-Landau's (2003: 41) concern is that idealising versions of constructivism fail to 'yield prescriptions that match up with our views about what constitutes paradigmatically moral and immoral behaviour'. Now, if this is true of *idealising* versions of constructivism, then it is likely to be even more of a concern when it comes to *non-idealising* views like Humean constructivism. This is because one of the key motivations for idealising the constructivist method in the first place is to try and achieve extensional adequacy with respect to paradigmatic cases of moral and immoral behaviour. If we do away with idealisation, then the probability of achieving extensional adequacy seems significantly lower.

Shafer-Landau (2003: 41–42) considers this failure to achieve extensional adequacy to be a huge problem for the constructivist. If the constructivist method

---

[20] In the next chapter, Chapter 4, I will present an idealising version *of* Humean constructivism, which I call 'enlightened constructivism'. Under enlightened constructivism, it is the hypothetical attitudes of hypothetical beings that go towards fixing truths about what moral reasons one possesses. In other words, I will attempt to modify Humean constructivism so that it becomes 'objectivist' rather than 'subjectivist' by the lights of Shafer-Landau's (2003: 41) taxonomy.

were to justify a policy such as segregation based on race, for instance, then we would intuitively feel that there was something wrong with the method. What this suggests is that our 'deepest settled convictions about core cases and principles must serve as constraints on moral theorizing' (ibid.: 42). To be clear, however, Shafer-Landau does *not* claim that our deepest settled moral convictions are self-certifying or self-evident. These convictions are not completely immovable. They are not set in stone like the Ten Commandments that God handed down to Moses. They are more like moral defaults that can be overturned on reflection and with very good reason.

As a supplementary point to this, Shafer-Landau adds that it is more plausible to start the process of moral theorising with some substantive views than with any particular hypothetical method. Indeed, it would seem that we cannot select a method in the first place without the aid of some substantive moral views. When constructivists defend their preferred version of the constructivist method they 'typically try to establish … [its] appeal by showing that both the initial conditions and the emerging principles match up closely with our deepest ethical convictions' (2003: 42). Though, of course, this is not the *only* way in which constructivists try to establish the appeal of their theories. They may also point to their theory's ability to account for the practical, action-guiding character of morality, say, or to the fact that it steers clear of controversial metaphysical assumptions and is compatible with a naturalistic worldview, etc.

This approach that Shafer-Landau (2003: 42) mentions – of trying to establish the appeal of a particular version of constructivism by showing how both the initial conditions of the method it employs and the truths which emerge from it correspond closely to our deepest settled moral convictions – raises the following dilemma for the constructivist. (Hereafter, when I speak of 'constructivists' or 'constructivism', I am referring to *Humean* constructivists and *Humean* constructivism specifically.) Do the initial conditions of the constructivist method incorporate moral constraints, or do they fly free of such constraints? The 'initial conditions' are here understood to be: (a) the full set of a given agent's evaluations, in either an occurrent or dispositional form; (b) any non-normative facts that are pertinent to those evaluations; (c) the bare-bones logical and instrumental requirements that govern practical reasoning, such as the principle of means-end coherence.

The question is, are these initial conditions constrained by our 'deepest settled convictions about core cases and principles', or are they unconstrained (Shafer-Landau 2003: 42)? The problem with the latter option is that there is no reason to expect that the truths which emerge from the (Humean) constructivist method will conserve our deepest ethical convictions. And here is where Shafer-Landau delivers a devastating blow against constructivism: 'To justify abandoning our central ethical views for those that would emerge from such a process [i.e. the constructivist method], we would need to have *greater warranted confidence in the process than in the particular views themselves* [emphasis added]. That is highly unlikely.' (ibid.).

The only alternative is for the constructivist to concede that the initial conditions of the constructivist method *do* incorporate moral constraints. But this would be to acknowledge that said moral constraints – i.e. certain moral reasons – are conceptually and explanatorily prior to the moral reasons that emerge from that method. These pre-existing moral reasons would not themselves be products of construction; they would obtain independently of constructive functions. And this is an implicit form of meta-ethical realism (Shafer-Landau 2003: 42).

This dilemma that Shafer-Landau sets out is reminiscent of Socrates' Euthyphro dilemma: 'Is the pious loved by the gods because it is pious, or is it pious because it is loved by the gods?' (*Euthyphro*, 10a). Socrates argued that the latter option seemed to make divine approval arbitrary, thereby undermining the goodness of the gods. Shafer-Landau's Euthyphronic dilemma for constructivism runs as follows. Do agents have moral reasons *because* there are moral reasons, or are moral reasons moral reasons *because* agents have them? In other words, is reason-giving status something that *precedes* the constructivist method or something that is *conferred* upon things/options in the world by it? For Shafer-Landau, the former is just realism in another guise, while the latter involves 'dignifying the arbitrary choices of … agents' (2003: 43).

3.2 A RESPONSE TO SHAFER-LANDAU'S OBJECTION

Shafer-Landau's objection deals a significant blow to constructivism. But is it a knock-out blow? Not quite. Constructivism is reeling, but it is not yet down and out for the count. The constructivist's response to Shafer-Landau's Euthyphronic dilemma is to bite the bullet and concede that the initial conditions of the

constructivist method are *not* constrained by our deepest settled convictions about core cases and principles. What this means, in practice, is that we cannot restrict the set of evaluations that are fed into the constructivist method on the basis of putative ethical truths – i.e. truths about reasons – that are thought to exist prior to the method.

What Shafer-Landau wants to say, translated into constructivist terms, is that certain evaluations have a kind of special status: they are settled and entrenched; the jury is no longer out on them; they are like ancient trees whose roots have burrowed deep into the soil, and it would take some extraordinary wind to uproot them. An example might be the moral judgment that the enslavement of blacks in the antebellum American South was a profound moral evil. Shafer-Landau's implicit view is that when *other* evaluative attitudes come into conflict with these *special* evaluative attitudes, those other attitudes must simply give way. So, for example, if an agent judges that slavery in the American South was actually a morally good thing, then this evaluation cannot be determinative – even in part – of ethical truth, because it is in stark conflict with the settled, non-negotiable evaluation that slavery was a profound moral evil. In short, we have the full set of human states of valuing, and within this full set there exists a subset that has a special, elevated status. This elevated subset places limits on which *other* states of valuing in the set are eligible to go towards fixing ethical truth.

This raises the question of how we can know which attitudes of valuing are settled, non-negotiable, not up for debate, etc., and which are not. Shafer-Landau does not go into much detail on this point. He merely says that our deepest settled moral convictions are *not* self-certifying or self-evident (2003: 42). Presumably, then, what makes an evaluation 'settled' is that it is: (a) held by the overwhelming majority of agents; (b) held with a high degree of confidence in its being correct.

Take, as an example, the judgment that the Holocaust was morally wrong. As Lenman (2000: 349) puts it, 'Here, if anywhere, surely, there is a considered moral judgment at stake that is well-enough entrenched not to be up for grabs in the cut and thrust of reflective equilibrium, a judgment far enough from the periphery of the web of our moral beliefs to furnish a compelling *reductio* of any theory that might undermine it'. And yet, under Shafer-Landau's view, the judgment that the Holocaust was morally wrong is *not* a self-certifying or self-evident moral truth. It is not, for

instance, a conceptual truth that is true in virtue of the essences of its constituent concepts; one need not be conceptually deficient to reject the judgment that the Holocaust was morally wrong. The judgment is merely an evaluation that is very widely held and in which we place an extremely high degree of confidence.

The relevance of all this is as follows. Shafer-Landau's main gripe with (Humean) constructivism is that if we allow the actual attitudes of agents to go towards determining ethical truth, then whatever truths emerge from the constructivist method are unlikely to conserve our deepest settled convictions about core cases and principles. This is perceived by Shafer-Landau (2003: 42) to be a problem because he thinks that we have more confidence in our bedrock moral convictions than in any constructivist method. However, under constructivism, our deepest settled convictions just are attitudes of valuing. And what makes these convictions 'deep' and 'settled' is that they are: (a) very widely affirmed by actual agents; (b) affirmed with a high degree of confidence on the part of said agents.

So, the fact of a moral conviction's being 'deep' and 'settled' depends on the contingently held evaluations of valuing creatures. But this is the *same kind of contingency* that the constructivist asks us to accept with respect to the deliverances of the constructivist method. Whether or not something is ethically true depends, at least in part, on the contingently held evaluations of valuing creatures. Shafer-Landau considers this kind of contingency to be problematic when it comes to determining truths about what moral reasons one possesses, yet he appears to implicitly accept it when it comes to determining which of our moral convictions are deep, settled, non-negotiable, not up for debate, etc. Perhaps Shafer-Landau has in mind some method that would allow us to determine which of our moral judgments qualify as deep, settled, etc., in a way that does *not* depend on the contingently held evaluations of agents. But, to my knowledge, he has not provided such an account.

To be clear, I am not claiming that Shafer-Landau's view on this point is self-contradictory or untenable. One *could* maintain that: (a) when it comes to determining which of our moral convictions are deep, settled, non-negotiable, not up for debate, etc., it is acceptable to rely on the contingently held, actual evaluations of agents; (b) when it comes to determining the nature of ethical truth, it is *unacceptable* to rely on the contingently held, actual evaluations of agents. I am merely pointing out this asymmetry.

As an example of one of our deepest settled convictions, Shafer-Landau (2003: 42) offers the judgment that 'a policy that entirely discounted the needs of the vulnerable' would be morally wrong. Now, the question of what makes this judgment one of our bedrock moral convictions is different to the question of what makes this judgment *true*. Shafer-Landau is (famously) a moral realist. He has a different view of what the truth-makers of moral judgments are than constructivists. But that is not what is at issue here.

What is at issue is how we know which of our substantive moral views are so deep and settled that they should serve as constraints on which evaluations can go towards determining the nature of ethical truth. Again, as far as I can tell, whether or not a particular moral judgment counts as being deep and settled seems to depend, ultimately, on the actual, contingently held evaluations of agents. If Shafer-Landau is willing to accept this sort of contingency in one domain, then perhaps this same sort of contingency is not as revisionist as it might seem when applied to the domain of determining truths about one's moral reasons.

There is a further reason for thinking that the constructivist view of things is less alarming and revisionist than it might, at first glance, seem. If our deepest settled convictions about core cases and principles *really are* our deepest and most widely-shared states of valuing, then this will be reflected in the moral reasons that are generated by the constructivist method. Consider the following analogy. Under constructivism, the role of actual states of valuing in constructing ethical truth is loosely analogous to the role of voting by citizens in electing members of parliament under a system of proportional representation. Each evaluation that an agent forms is like a vote. It is a vote to determine the nature of ethical truth.

Imagine that 99 per cent of the agents living in a moral community affirm – i.e. vote for – a particular moral judgment. Let us say that the judgment in question is that beating one's children is morally wrong. If 99 per cent of agents affirm this judgment, then for that 99 per cent of agents, this judgment will play a role in determining what moral reasons they possess. It seems probable that, under constructivism, virtually all of the agents who endorse this moral judgment will have a reason – though not necessarily an overriding one – to refrain from beating their children.

I say 'almost all' rather than 'all' because we can easily imagine cases where there are confounding variables. For instance, an agent might live in a household where his children are sociopathic bullies who are constantly trying to beat *him* up. Depending on this agent's other evaluations and on what the relevant non-normative facts are, this agent may have a reason to give his children a good thrashing as a pre-emptive measure to stop them from giving him one. Alternatively, you could have a situation where, for whatever bizarre reason, an agent is somehow able to save the lives of thousands of innocent people only if she immediately beats up her children.

The point is that insofar as our deepest settled moral convictions *really are* our strongest and most widely-shared evaluative attitudes, we can expect them to exert an outsize influence on determining what reasons any given agent possesses. To return to the voting analogy, we can think of the cluster of our deepest settled moral convictions as being a little bit like a popular political party that many, many people vote for. Because so many agents 'vote' for our deepest settled moral convictions – e.g. because so many agents sincerely affirm that the Holocaust was a profound moral evil – these bedrock convictions win a large number of seats in our moral parliament and play a correspondingly large role in fixing ethical truth. For this reason, we actually should *expect* the moral reasons generated by the constructivist method to conserve most of our bedrock moral intuitions regarding core cases and principles.

Constructivism is a very democratic meta-ethical theory insofar as it outsources the determination of ethical truth to the 'votes' – i.e. *actual* evaluative attitudes – of individual moral agents. The constructivist's motto is to 'let the people decide' what the nature of ethical truth should be rather than leaving it up to some privileged elite. By 'privileged elite', I mean, for example, those hypothetical beings under an ideal observer theory such as Firth's (1952) whose hypothetical attitudes go towards determining the nature of ethical truth. Under such views, *actual* agents are deemed to be insufficiently informed, imaginative, coherent, dispassionate, etc., to be trusted with the right to 'vote' (vote to determine the nature of ethical truth, that is). The thought is that we must restrict the franchise, as it were, to ideal beings who satisfy certain criteria. Firth (ibid.: 335), for instance, proposes that such beings must be 'omnipercipient' in that they must be able to simultaneously 'visualize all actual facts,

and the consequences of all possible acts in any given situation, just as vividly as …
[they] would if … [they] were actually perceiving them all'; they must also be
dispassionate in that they are 'incapable of experiencing … such emotions as
jealousy, self-love, personal hatred, and others which are directed towards particular
individuals as such' (ibid.: 340).

To bring this back to Shafer-Landau's (2003: 42) objection to constructivism,
Shafer-Landau's view, when translated into the terms of my voting analogy, is that
there is an established political orthodoxy – i.e. our deepest settled moral convictions
– that should constrain *which* votes can be counted in the democratic process that is
the constructivist method. So, for instance, let us say that part of the established
political orthodoxy is that Nazism is a heinous, irredeemable ideology that simply has
no place in our contemporary politics. Shafer-Landau's view is that anyone who
votes in contravention of the political orthodoxy on this point – e.g. anyone who votes
for the Nazi party – should not have their vote counted. The constructivist view, on
the other hand, is that such people *should* still have their votes counted. Even 'bad'
votes are still legitimate votes. And we are free to outvote such people by voting for
more moderate, centrist parties.

As discussed earlier, one of our deepest settled moral convictions is that the
Holocaust was a profound moral evil. We can think of this judgment as forming part
of our established moral orthodoxy. Shafer-Landau's view is that heterodox
judgments on this point – e.g. the Holocaust was actually a profound moral *good* –
should not, to whatever degree, be allowed to determine what moral reasons agents
have. Such judgments should be excluded from the method that we use for
determining ethical truth. The constructivist takes a different line. Heterodox moral
judgments that are in radical opposition to our settled moral orthodoxy are still part of
the material that is used to construct ethical truths. Agents who endorse views that
are abhorrent to most members of the moral community, such as racists, neo-Nazis,
anti-Semites, etc., should still have their votes counted.

As mentioned earlier, this voting analogy is only a loose one. For the sake of
precision, let me note one of the ways in which it is inaccurate and potentially
misleading. Under constructivism, moral reasons are only ever reasons for a
particular agent. Reasons are always relativised to a particular evaluative standpoint
(Street 2008: 224). Each of our evaluations can be thought of as a vote, but it is a

vote to determine the nature of our *own* moral reasons, not those of other people. When we evaluate things/options in the world, we are not voting for a single, shared set of moral reasons that apply to all agents; we are voting merely for our own. There is no shared moral parliament, then, that governs all agents. It is more like each agent is governed by their own moral parliament and elects members – i.e. reasons – to it.

Having said this, the earlier point I made vis-à-vis Shafer-Landau's objection still stands: insofar as our deepest settled moral convictions *just are* the strongest and most widely-shared states of valuing among agents, we can expect them to exert an outsize influence on determining what reasons any particular agent possesses; we can expect those widely-shared and deeply-rooted evaluations to, other things being equal, generally translate into shared reasons among agents. On top of this, it is worth emphasising – as I did in §2.1 – that the link between what an agent's states of valuing are and what reasons they possess is indirect rather than direct. Under constructivism, it is the *full set* of an agent's evaluations that go towards determining truths about what reasons they possess, and agents can often fail to notice how certain evaluations are in tension with one another; they can fail to grasp the full implications of certain evaluations; they can fail to keep in view all of the evaluations that they do, in fact, hold, etc.

So, for example, if an agent sincerely judges that the Holocaust was actually a tremendous moral good, that does not automatically mean that they will have reason to go around trying to create the conditions for a second Holocaust. This moral judgment might be incompatible with other evaluations that they hold, such as the conviction that it is morally wrong to murder innocent people, that everyone has the right to a fair trial, that people should be judged as individuals rather than as members of groups, etc.

Moreover, there are a great many non-normative facts that would be highly relevant to this particular judgment about the Holocaust. This agent's judgment – that the Holocaust was a tremendous moral good – is more than likely predicated on certain factual falsehoods and historical inaccuracies. The agent may believe, for instance, that the Jews really were responsible for Germany's defeat in the First World War; and that it was the Jews who ensured that Germany was subsequently humiliated in the Treaty of Versailles; and that the Jews were lending money to the

impoverished German people at exorbitant rates of interest; and that the Jews were using their stolen wealth to bribe and corrupt German officials and politicians; and that the Jews were sexually exploiting German maidens, etc. Needless to say, such beliefs would be completely delusional and ahistorical. And that matters, because *facts* matter in determining truths about what reasons one possesses.

3.3 Understanding the Amoralist Threat to Constructivism

Let us now move on to the second key objection to constructivism: the amoralist objection. In short, the amoralist objection is that if constructivism is true, then hypothetical amoralists are possible, and the mere possibility of such creatures is a profoundly counter-intuitive result for a meta-ethical theory to generate, a result that makes that particular theory less plausible. Broadly speaking, a hypothetical amoralist is an agent who lacks any reasons to be moral.

We can imagine, for instance, a hypothetical version of the sadistic Roman Emperor Caligula who *correctly* judges himself to have no moral reasons to do anything at all (Hopster 2017: 774). As in, what we are imagining here is that no moral reasons are entailed from within his evaluative standpoint (i.e. the full set of his attitudes of valuing – in either an occurrent or dispositional form – together with any non-normative facts that have a bearing on those attitudinal valuations). Let us say that this hypothetical Caligula delights in flaying people alive and then feeding them to his hunting hounds. We can imagine him saying something like this:

> Look, I know that torturing innocent people for one's own amusement is morally wrong – morally *evil*, in fact – but I couldn't care less. I lack all trace of moral feeling. It's not that I don't *understand* morality. On the contrary, I have a highly-developed understanding – all the better that I can flout its norms! But morality does not bind me. It is not a tune that I march to. What I seek out of life is to torture and terrorise the people of Rome, to keep them in a constant state of fear and anxiety. No act is too perverse or debased for me. I am a scourge come to whip mankind, an emissary of Hades.

The worry here is that an agent of this kind is possible under constructivism, and this seems to violate our intuition that all agents possess at least some significant – but

not necessarily overriding – reasons to be moral, such as reasons not to torture innocent people for one's own amusement.

At this point, I think it is important to distinguish between two different kinds of amoralist. This is not a distinction that I have seen drawn anywhere in the literature on this topic, but it is, I believe, an important and clarifying distinction to make. When we speak of hypothetical amoralists, there are actually two sorts of agent that we might have in mind. The first are those who are perfectly indifferent to and unmoved by moral considerations. It is not that these agents lack an *understanding* of morality. On the contrary, these agents know full well the sorts of reasons they *would* have if they cared about morality. But these agents do not care a whit about morality or living moral lives at all. That is to say, they do not make any *sincere* moral judgments, and consequently, no moral reasons are entailed from within their evaluative standpoints.

As Street (2012: 55–56) puts it, '[Under constructivism], moral feeling is something that an agent ultimately either has or does not have as part of his nature, and if an agent has no trace of it anywhere in his evaluative makeup, or if that feeling does not run deep enough when push comes to shove, then this is not a mistake in any genuine sense ... if one lacks moral concerns altogether, then morality does not bind one.' I will refer to this particular form of hypothetical amoralism – i.e. when an agent makes no sincere moral judgments and has no moral reasons – as 'Amoralism of Indifference'.

Before I outline the second kind of hypothetical amoralism, let me clarify one terminological point. In using the qualifying term 'sincere' when talking about moral judgments, I am referencing a distinction that I drew in §2.2. To recap, a moral judgment is sincere if it is a genuine, unfeigned evaluative attitude on the part of the agent who holds it, as opposed to one that is held with some degree of dissimulation or pretence. An example of dissimulation or pretence would be if an agent is making a moral judgment in an 'inverted-commas' sense (Hare 1952: 124). That is, making a judgment with *sociological* content rather than *normative* content.

I think that this inverted-commas interpretation is the most plausible reading of Emperor Caligula's 'moral' judgments in my little vignette above. When Caligula says, for instance, that 'torturing innocent people for one's own amusement is morally wrong', this judgment amounts to a *descriptive* claim rather than a *normative*

one. The judgment should be read in a 'non-evaluative, inverted-commas sense, as meaning that so-and-so falls within a class of actions which is generally held (but not by him) to be obligatory in the evaluative, imperative-entailing sense' (Hare 1952: 165). Caligula takes an almost anthropological interest in morality: it is a system that he has studied with keen interest so that he can learn how to better flout its norms. When he makes a moral judgment, he does so with a sense of wry detachment; he views himself as being *above* morality, as having transcended it.

The second kind of hypothetical amoralism is as follows. We can imagine agents who possess moral reasons that seem utterly perverse and deranged from our perspective. It is not that these agents are *indifferent* to morality: they *do* make sincere moral judgments. But the moral judgments they make are perverse – or, at least, they seem so from our perspective. And this perversity in their moral judgments leads (indirectly and depending on the relevant non-normative facts) to their having perverse moral reasons. A quote from the classic novel *Madame Bovary* comes to mind: 'She [Madame Bovary] was beginning now and then to express peculiar opinions, condemning what everyone else approved and approving things that were perverse or immoral – a way of talking that made her husband stare at her wide-eyed' (Flaubert 1992: 77).

We can imagine a hypothetical version of Madame Bovary forming a whole series of moral judgments that make other agents stare at her wide-eyed. In the novel *Madame Bovary*, the character Madame Bovary engages in a string of adulterous liaisons. Her husband, Charles, is blissfully unaware that his wife is making a cuckold and a fool out of him. We can imagine a hypothetical Madame Bovary endorsing such judgments as: 'It is morally good to have adulterous affairs. In fact, the amount of good scales in a linear way with the number of affairs one has. The more affairs one has throughout one's marriage, the more good one is doing. What is morally *bad* is to remain faithful to one's spouse. The worst thing one can do is to go one's whole marriage without having any affairs at all.'

There is no *indifference* to morality on the part of Madame Bovary. Rather, it is that she is operating with an idiosyncratic moral code that diverges markedly from our own. Indeed, she may consider herself to be a thoroughly ethical person. I will refer to the kind of hypothetical amoralism that she exhibits as 'Amoralism of Divergence'. To be clear, what is objectionable here is that constructivism allows that

there is a possible world in which it is entailed from within Madame Bovary's evaluative standpoint that she has a moral reason to engage in adulterous affairs, say. Essentially, this is a problem of an agent having *perverse* moral reasons (as opposed to their having *no* moral reasons at all).

Given that the Amoralism of *Indifference* objection and Amoralism of *Divergence* objection are distinct, it makes sense to treat them separately. In §3.4 I will evaluate the former; in §3.5 I will evaluate the latter.

3.4 HYPOTHETICAL AMORALISM OF INDIFFERENCE

I will endeavour to show in this section that constructivism is more than capable of parrying the Amoralism of Indifference objection. Let me begin by offering an example of a hypothetical being who exhibits this form of amoralism. In the *Canterbury Tales* by Geoffrey Chaucer (1985) there is a very interesting and peculiar character known as 'the Pardoner'. In medieval England, pardoners were people who bore an 'official letter from the Pope allowing … [them] to grant remission of time in Purgatory for those who agreed to do fixed sorts and amounts of penance' (ibid.: 485–486). Consider the following passage from the *Canterbury Tales*, in which the Pardoner describes his approach to his profession:

> I take pains to preach
> In churches with a lofty, resonant voice,
> Regular as a bell I ring it out,
> For everything I say I have by heart ...
> Then I bring out long boxes made of glass,
> Chockful of rags and bones, each one a relic,
> Or so they think ...
> And by this dodge I've gained, year after year,
> A hundred marks since I was pardoner [about £350,000 in today's money] ...
> I stand up in my pulpit like a priest;
> When the bumpkins have all settled in their seats
> I preach just as you've heard me say; and tell
> A hundred taradiddles more as well.
> And then I do my best to stretch my neck

And bob my head in every direction

Over the folk, now this way, and now that,

Just like a pigeon sitting on a barn.

Both hands and tongue are busy, and so quick

That it's a joy to watch me at my work.

I preach only of avarice and the like,

And in this way induce them to be free

In giving cash – especially to me.

Because my only interest is in gain;

I've none whatever in rebuking sin.

No, none! When they are pushing up the daisies,

Their souls, for all I care, can go to blazes ...

I preach for money, and for nothing else ...

Thus I know how to preach against the vice

Which masters me – and that is avarice.

Though I myself am guilty of the sin,

I know how to make other people turn

From avarice, and bitterly repent ...

You'll never catch me working with my hands –

At begging I can make a better living

Than St Paul ever did at basket-making;

He's an apostle I won't imitate.

For I'll have money, wool, and cheese, and wheat,

Though given by the poorest serving-lad,

Or by the poorest widow in the place,

Were all her children dying of famine.

No, no! I'll drink the ichor of the vine,

And have a pretty girl in every town ...

For although I'm a pretty vicious chap,

I can tell a story with a moral to it:

Here's one I preach to bring the money in.

Now if you'll all be quiet, I'll begin (Wright 1985: 325–328).

So, the Pardoner is a classic grifter and knave, a kind of travelling salesman blessed with the gift of the gab. Given that part of his job involves preaching to the 'bumpkins' about various sinful behaviours, it seems safe to assume that the Pardoner is quite knowledgeable about moral matters. He would be able to engage in the kinds of 'prolonged and intelligent conversations about moral matters' that make one a competent user of moral language (Svavarsdottir 1999: 188).

The Pardoner is aware of the kinds of considerations that *would* apply to him if he were the sort of person who cared about morality and living a moral life. He knows, for instance, that it is morally wrong to fleece the 'poorest widow in the place', even those whose 'children [are] dying of famine' (Wright 1985: 328). He knows this, and he *does not care*. He does not give a fig for morality. His evaluations do not aim at living a moral life. As Giuseppe Tomasi di Lampedusa (2007: 102) puts it in his classic novel *The Leopard*: '[F]ree as he was from the shackles imposed on many other men by honesty, decency, and plain good manners, he moved through the forest of life with the confidence of an elephant which advances in a straight line, rooting up trees and trampling down lairs, without even noticing scratches of thorns and moans from the crushed.'

We can imagine a hypothetical version of the Pardoner who correctly judges himself to have no moral reasons. That is, he makes no sincere moral judgments, and consequently, no moral reasons are entailed from within his evaluative standpoint (this follows because, under constructivism, a necessary condition for having moral reasons is that one holds sincere moral judgments). Any moral judgments that the Pardoner appears to make, such as when he is preaching to the masses, are made purely for theatrical effect: they are part of his performance; they are judgments with sociological rather than normative content.

The objection here is that the mere possibility that a character like the Pardoner could be *correct* in judging himself to have no moral reasons runs 'counter to common-sense ... our ordinary thinking seems to support the idea that some normative judgments – typically *moral* judgments – are objectively binding in the sense that *all* agents should be responsive to them' (Hopster 2017: 774). The thought here is that morality does not have an opt-out clause: one cannot simply forfeit all moral reasons in virtue of being totally indifferent to moral concerns, in virtue of having an idiosyncratic evaluative standpoint. As Foot (2002: 162) puts it,

'morality is supposed to be inescapable in some special way … No one, it is said, escapes the requirements of ethics by having or not having particular [evaluations]'. We think – i.e. evaluate – that it is morally wrong for the Pardoner to sell people fake relics and to swindle money out of those living in penury; we judge that he has moral reason to stop defrauding and fleecing people. And it seems counter-intuitive to say that he can escape the relevant requirements here in virtue of lacking all moral feeling.

What does the constructivist have to say in response to this? The first point to make is that in the actual world or in any world approximating our own, a person like the Pardoner would be 'in all likelihood inconsistent, wrong about many of the non-normative facts, and self-deceived' (Street 2009: 294). The Pardoner makes his living by travelling from village to village selling phoney relics to poor countryfolk. In any world similar to that of medieval Europe, one could probably get away with a con of this sort for a while, but there would be a high degree of risk involved. One can only fleece people for so long before they start cottoning on to the fact. If enough people start to realise that the 'relics' they were sold are actually just bits of old rag and bone, then word would spread, and before he knew it the Pardoner might find himself lynched by an angry mob armed with pitchforks (who would, presumably, exact some horrible medieval punishment on him in reprisal). The point is that in any world approximating our own, the Pardoner could easily be *mistaken* in thinking that he has no reason to stop swindling people.

One might worry that I have skirted the issue here by blurring the distinction between moral and normative reasons. I have pointed out that in any world approximating our own, the Pardoner would, due to the relevant non-normative facts, most likely have reasons to stop swindling people (e.g. because of the risk involved and the potential repercussions if he is caught). But we intuitively feel – or so the objection goes – that the Pardoner has *moral* reasons to stop swindling people, not just *normative* reasons relating to his own self-interest.

This is a tricky issue. Recall from §1.3 that constructivism locates the moral within the domain of the normative, so an agent's moral reasons are just a subset of the full set of their normative practical reasons more generally. Normative truths – i.e. truths about one's normative reasons – are constructions from the attitude of valuing, and

so are ethical truths. There is no fundamental difference between moral reasons and normative reasons. They are not 'normatively independent' and do not belong to 'non-overlapping parts of the normative domain' (Forcehimes and Semrau 2018: 700).

What the distinction between moral and normative reasons boils down to, under constructivism, is a general – though far from exceptionless – difference in their substantive content. Moral reasons tend to be other-regarding in character. One's moral reasons generally concern the 'interests or welfare either of society as a whole or at least of persons other than the judge or agent' (Haidt 2003: 853). They concern the well-being and flourishing of conscious creatures *other* than the agent who possesses them. *Non*-moral normative reasons, by contrast, tend to be self-regarding in character: they tend to be concerned with one's own interests, well-being, flourishing, etc.

So, moral reasons just are normative reasons whose content is other-regarding in character. In order to have moral reasons under a constructivist view of things, one must make at least *some* judgments that concern the 'interests or welfare either of society as a whole or at least of persons other than the judge or agent' (Haidt 2003: 853). The upshot of this is that in order for a hypothetical version of the Pardoner to have literally no moral reasons, he would have to make *no evaluations with other-regarding content at all*. This is a very stringent criterion to meet.

We are social primates who evolved in a context of small, tight-knit communities that involved a high degree of mutual dependence; because of this, our own well-being is intimately bound up with the well-being of others (Pinker 2022: 68). In any world even vaguely approximating our own, the Pardoner would surely have at least *some* judgments concerning the interests or welfare of agents other than himself. Even if he doesn't care about the well-being of the country bumpkins whom he is duping, he surely cares – at least to some degree – about his own parents, children, close friends, etc.

As soon as we grant that the Pardoner has at least some evaluations with other-regarding content in his motivational set, this opens the door to his: (a) having moral reasons that are entailed from within his own evaluative standpoint; (b) his being mistaken as to what those reasons are (e.g. through his being inconsistent or self-deceived in various ways). For instance, the Pardoner might implicitly judge that it

would be wrong for a travelling huckster to swindle his (the Pardoner's) parents out of their life savings; he might judge this without ever putting two and two together and realising that this is precisely the sort of thing that he is doing to other people's agèd parents. (To be clear, I am not claiming that there is any kind of logical contradiction involved in this, but merely that the Pardoner might not fully appreciate the implications of certain evaluations that he holds.)

At this point, I think a critic of constructivism would probably say something like, 'Okay, so you've shown that it is hard to imagine an agent – at least in any world approximating our own – making *literally* no evaluations with other-regarding content. And you've shown that if an agent evaluates regarding others, then there are various ways, under constructivism, in which they can be mistaken about what moral reasons they possess (e.g. by not knowing some of the non-normative facts that have a bearing on their evaluations). But given that we're talking about *hypothetical* agents here, why can't we just *stipulate* that they make no sincere moral judgments at all and that, furthermore, they are ideally coherent and informed? So, we can imagine a hypothetical version of the Pardoner who holds no evaluations with other-regarding content at all – in either an occurrent or dispositional form – and who is not suffering from any failures of coherence or ignorance concerning the non-normative facts. Such an agent would, under constructivism, *correctly* judge himself to have no moral reasons at all, including any moral reasons to stop swindling and fleecing people, and therein lies the problem.'

The constructivist retort to this is as follows. Ideal coherence is a very high threshold to meet (Hopster 2017: 775). And any Amoralists of Indifference, such as the Pardoner, who met it would be more akin to alien life-forms than human beings (Street 2009: 281). And when it comes to alien life-forms, our intuitions shift (Sobel 2016: 34). First, let us take a moment to drill down on what exactly we are stipulating here. We are stipulating that, in the case of the Pardoner, this is a person who: (a) has all of the relevant non-normative facts in hand, including understanding the likely consequences of his actions on the lives of others, understanding all of the alternative ways in which he could be maximising his well-being and flourishing in life, etc.; (b) grasps the full implications of all of his evaluations; (c) is able to keep in view all of the evaluations that he does, in fact, hold; (d) does not have any *logical* contradictions in the full set of his evaluations (as discussed in §2.1, the

constructivist method for generating truths about one's moral reasons does not eliminate *practical* contradictions from an agent's motivational set); (e) *makes absolutely no sincere moral judgments* (since, if he did, he would probably have moral reasons that stem from those judgments).

Of course, such a person is *conceivable*: we can conceive of a hypothetical version of the Pardoner who meets all of these criteria and who then correctly judges himself to have no moral reason whatsoever to stop defrauding and fleecing people. But such a being would be a very strange specimen indeed; he would possess a psychology that differs radically and bizarrely from anything that we are familiar with. Such a being would look 'more like an interesting visitor from another planet than a human being' (Street 2009: 281).

The empirical evidence on this topic suggests that '[w]here aliens are concerned, ordinary respondents do *not* [emphasis added] think that objective moral judgments retain their binding authority' (Hopster 2017: 775). The data tentatively suggest that most people are actually '*graded* objectivists' (ibid.). That is, they have absolutist intuitions when moral judgments concern individuals from their own culture – i.e. they do not think that individuals from their own culture can escape moral requirements in virtue of having an idiosyncratic evaluative standpoint – but their intuitions take a 'strikingly relativist turn when they are encouraged to consider individuals from radically different cultures or ways of life' – the extreme case being aliens who inhabit a different galaxy (Sarkissian 2011: 500). When it comes to aliens, our intuitions shift such that we do not feel that aliens are bound by the moral requirements that we think bind human beings (Sobel 2016: 34).

Let me offer an example to make this a bit more concrete. Consider the following extract from Richard Dawkins' (1996) book *Climbing Mount Improbable*:

> [Female spiders] are larger and more prominent [than male spiders]. Couple the larger size of females with the fact that spiders, of any age or sex, tend to eat anything smaller than themselves that moves, and it does raise problems for males … Any spider who ventures on to the web of a larger spider is in mortal danger, but this is a danger that a male must face if he is to do what he has to do. Exactly how the male copes with the problem varies from species to species. In some cases he wraps a fly in a silken parcel and presents it to

the female. He waits until her fangs are safely sunk in the fly before he goes to work on her with his sexual apparatus. Males without a fly parcel may be eaten (ibid.: 40–41).

Imagine an alien species of giant, intelligent, self-aware spiders existing in the Andromeda Galaxy. These alien spiders resemble Earthly spiders in outward form, but they are much larger. The males of the species are about the size of a Labrador Retriever, while the females are roughly the size of a grizzly bear. These alien spiders practise the same kinds of mating rituals described in the above passage from Dawkins. After copulating with the male spiders, the female spiders often proceed to murder them and eat them.[21]

It is morally wrong to murder and then eat one's romantic partner. I think it is fair to say that we have absolutist intuitions regarding this moral judgment when it comes to individuals from our own culture. But the same is not true when it comes to the example of the alien spiders. We may be disgusted by the mating rituals of the alien spiders. We may be horrified by the prospect of ever ending up on their planet. But it does not seem intuitive to say that the female spiders are *morally wrong* to kill and eat the male spiders or that they should be responsive to moral considerations relating to this action.

So, when it comes to alien life-forms, our moral intuitions relating to the inescapability of certain moral requirements flip from absolutist to relativist. In order for an agent to be an ideally coherent amoralist, they would 'have to be an alien being – if not literally, then something very close to it' (Street 2009: 281). So, the mere possibility of hypothetical amoralists (of indifference) under constructivism does not, in fact, violate our pre-theoretical moral intuitions in this area. Indeed, the fact that constructivism allows for this sort of moral diversity across galaxies appears to *harmonise* with our intuitions.

---

[21] In the case of these aliens we can, but need not, stipulate that they are ideally coherent and informed; our intuition that they are *not* making a genuine moral mistake – i.e. a mistake about the moral facts – is clear enough even without this stipulation.

3.5 HYPOTHETICAL AMORALISM OF DIVERGENCE

This seems like a good moment to segue into a discussion of the second kind of amoralism that I sketched out earlier: Amoralism of Divergence. Recall that hypothetical amoralists of this flavour are not *indifferent* to morality; they do not lack moral concerns in their evaluative makeup (as in the case of the Pardoner). Rather, it is that they make sincere moral judgments and possess moral reasons that seem utterly perverse and deranged from our perspective.

I gave the example earlier of a hypothetical version of the character Madame Bovary from Flaubert's classic novel *Madame Bovary*. In the novel, Madame Bovary cuckolds her husband by engaging in a string of adulterous liaisons. We can imagine a hypothetical version of Madame Bovary sincerely making the following judgments. 'It is morally good to have illicit affairs. In fact, the amount of good scales in a linear way with the number of affairs one has. The more affairs one has throughout one's marriage, the more good one is doing. What is morally *bad* is to remain faithful to one's spouse. The worst thing one can do is to go one's whole marriage without having any affairs at all.'

The objection here is that constructivism allows for the possibility of a world in which it is entailed from within Madame Bovary's evaluative standpoint that she has a *moral* reason to engage in adulterous affairs; hypothetical Madame Bovary would be *correct* in judging herself to have a moral reason to do so – and this seems like a counter-intuitive result for a meta-ethical theory to generate.

This objection generalises beyond the case of adultery. It is generally thought, for instance, that murdering innocent people is morally wrong. But constructivism seems to allow for the possibility of agents who, due to their perverse evaluative standpoints, actually have moral reasons *to* go around murdering innocent people, such as by sneaking up behind people and shoving them down a crevasse. (Recall that for the constructivist, a moral reason is just a species of normative reason whose substantive content is other-regarding in character, in the sense of being concerned with the well-being and flourishing of conscious creatures other than the agent who possesses them.) The objection here is that the bare possibility of such agents runs afoul of our core moral intuitions.

My response to the Amoralism of Divergence objection is very similar to my response to the Amoralism of Indifference objection. Let me stick, for now, with the example of hypothetical Madame Bovary. The first point to make is that in the actual world or in any world resembling our own, to continually engage in illicit affairs throughout one's marriage would require one to spin an elaborate web of lies for years on end. One's whole life would become a kind of lie. And if one is ever caught, it could well spell the end of one's marriage. Think of how bizarre an agent's motivational set would have to be for there to be no issues of coherence between judging that committing adultery is one of the highest goods and the rest of one's evaluations. In any psychology approximating our own, living a double life as a serial adulterer would carry with it a whole host of opportunity costs and negative implications. Madame Bovary would have to use up a good deal of her cognitive bandwidth simply keeping track of all of her fabrications. Famously, the novel *Madame Bovary* does not have a happy ending: it ends with Madame Bovary committing suicide after the years of duplicity finally catch up with her.

Of course, we can, as in the case of the Pardoner, simply stipulate that a hypothetical Madame Bovary is ideally coherent and informed. By making this stipulation, we can wish away any failures of coherence or ignorance concerning the non-normative facts that Madame Bovary might be suffering from. In doing this, we preclude the constructivist from pointing to these failures of coherence as a way of showing that perverse moral reasons do not, in fact, follow from Madame Bovary's evaluative standpoint – e.g. the constructivist cannot say that Madame Bovary is wrong in thinking that she has a moral reason to commit adultery since she does not fully comprehend the ramifications of her actions both for herself and for others.

We can stipulate that hypothetical Madame Bovary: (a) has all of the relevant non-normative facts in hand, including understanding the likely consequences of her actions on the lives of others, understanding all of the alternative ways in which she could be maximising her well-being and flourishing in life, etc.; (b) grasps the full implications of all of her evaluations; (c) is able to keep in view all of the evaluations that she does, in fact, hold; (d) does not have any logical contradictions in the full set of her evaluations.

But the problem with making these various stipulations is that when we take the time to carefully and accurately imagine what such a being would be like, we find

that they would be more akin to an alien life-form from another galaxy than a woman from 19th century provincial France. And, as discussed in the preceding section, when it comes to alien life-forms, our intuitions relating to the inescapability of certain moral requirements flip from absolutist to relativist (Hopster 2017: 775). To bring this into clearer relief, let me offer one more alien example. This time, it will *not* involve aliens who escape certain moral requirements that we intuitively think apply to all human beings (e.g. reasons not to kill and then eat one's romantic partner). Rather, it will involve aliens who possess moral reasons that seem *perverse* from the point of view of human beings. In short, it is a problem not of escaping moral reasons but of having radically divergent moral reasons.

Imagine that members of a highly advanced alien civilisation one day land a spaceship on Earth. They are 500,000 years older than Homo sapiens and are in possession of incomprehensibly sophisticated technology (allowing for faster-than-light travel, etc.). They are massive in size, squid-like in form, and enjoy a vastly richer subjective experience than human beings. Cognitively, they stand in relation to human beings as human beings do to tadpoles.

These alien travellers have landed on Earth because the anti-matter generator in their spaceship is in need of urgent repairs; it has been damaged during interstellar travel. The repairs require a particular metal that is found in abundance in the Earth's molten core but which is scarce elsewhere in space. The aliens insert a giant drill into the Earth's crust and use it to burrow down through layers of rock and magma to the molten core. They then use suction to extract part of the molten core and store it safely in the cellars of their spaceship. Given that the Earth's molten core is what sustains the protective magnetic field around the planet, and given that the magnetic field is what blocks harmful efflorescences of solar radiation, this action by the aliens renders the Earth uninhabitable and leads to the death of billions. As the ones being irradiated in this scenario, we would no doubt condemn this extractive action by the aliens as a heinous moral evil.

But the aliens view things rather differently. They judge that what they are doing is actually a great moral good because it is allowing them to repair their failing spaceship and return to their own galaxy (let us imagine that they have been exploring and charting neighbouring galaxies). The aliens bear no malice towards us, but because they stand in relation to us as we do to tadpoles they simply view our

destruction as an unfortunate side-effect of an otherwise good and necessary action: drilling into the Earth's core. By analogy, if a human mining operation had to drain a lake teeming with tadpoles in order to extract some extremely valuable and scarce minerals buried beneath the lake bed, there would be no hesitation over the suffering caused for the tadpoles; this is true even if the miners in question bore no malice towards the tadpoles.

In this example, the aliens *sincerely* judge that what they are doing is morally right. Let us say that it is entailed from their evaluative standpoints that they have moral reasons to drill down to the Earth's core and, as an unintended consequence, destroy the Earth's protective magnetic field. Does it violate our moral intuitions to suppose that these super-intelligent, ultra-sentient, squid-like aliens *really do* have moral reasons to perform this action?

The constructivist would insist that the answer here is 'no'. Intuitively, we do not feel that the aliens are making a genuine mistake about the moral facts in this scenario. Of course, we still find their moral worldview perverse, terrifying, and scarcely comprehensible. But we would not be inclined to say that they are wrong about the moral requirements that they think they have or that they ought to be constrained by the moral requirements that we think ought to constrain human beings.

In sum, when we take the time to carefully and accurately imagine what ideally coherent and informed hypothetical amoralists would be like, we find that they are more like aliens from another galaxy than flesh-and-blood human beings. And when it comes to aliens from another galaxy, it does not, in fact, seem counter-intuitive to suppose that such creatures might: (a) not share the moral reasons that we think all or most human beings share (i.e. exhibit Amoralism of Indifference); (b) possess their own, alien moral reasons that seem perverse from our standpoint but which are valid from their standpoint (i.e. exhibit Amoralism of Divergence). So, in allowing for the possibility of hypothetical amoralists, constructivism does not seem to run afoul of our pre-theoretical moral intuitions in this area; the theory is not revisionist in this respect.

3.6 THE RADICAL CONTINGENCY OBJECTION TO CONSTRUCTIVISM

As I discussed in §1.1, (Humean) constructivists deny that, in virtue of the structure of practical rationality and practical agency, all agents *necessarily* share a set of moral reasons – or any other normative practical reasons. The bare-bones standpoint of pure practical reason does not commit one to any specific substantive values (Street 2010: 370). It is compatible with an agent's being a valuer that they hold any substantive evaluation at all. The upshot of this is that under constructivism, any convergence in the normative practical reasons of agents is purely contingent. There is no guarantee, in advance, that any particular agent will have any particular reason whatsoever. We can only figure out what an agent's reasons are once we know what they actually value.

Dale Dorsey (2018) sees this sort of contingency as objectionable – though he ultimately goes on to defend his own version of constructivism (ibid.: 585ff.). According to Dorsey (ibid.: 577–578), constructivism 'denies that there are any reasons that are generally shared', and while we 'certainly wouldn't want to hold that all agents share all reasons, it would appear odd to hold that there are no reasons that we all share'. Dorsey gives the following example of a reason that we all share: 'the fact that children need minimal care to survive and flourish is a reason for anyone to provide such care for their (or indeed anyone's) children. But [constructivism] denies this: if a person doesn't care for her children or does not include her children's well-being as part of her evaluative makeup … failing to provide such minimal care "is not a mistake in any genuine sense"' (ibid.: 578).

Dorsey's background assumption here is that normative practical reasons are those facts that *count in favour* of particular actions; they are considerations that favour action; they are facts that 'can be said' for an agent acting in certain ways (Dancy 2006: 45; Finlay 2006: 1; Scanlon 1998: 17). I am not disputing this background assumption; I am merely making it explicit as it is relevant to understanding the radical contingency objection. To reduce the need to commit to any controversial assumptions regarding the metaphysical nature of reasons, we can interpret talk of reasons as referring to *correct ascriptions* of reasons rather than to *reasons themselves* (i.e. mysterious entities existing 'out there' in the ether).

So, Dorsey's (2018: 578) claim is that under constructivism, there are no reasons that *all* agents share (I will argue in §3.7 that this claim is false and based on a

misunderstanding of constructivism, but for now, I will just focus on unpacking Dorsey's objection). What is objectionable about this, according to Dorsey, is that it runs afoul of our intuitions. Intuitively, we think that there *are* reasons that all agents share, such as reasons to care for and nurture one's offspring (ibid.). But if an agent lacks the relevant evaluations, then, under constructivism, he cannot be said to be making a genuine normative mistake if he fails to possess reasons that *we* think he should possess.

Dorsey (2018: 578) goes on to say that it would be 'very strange' if everything that counted in favour of a given action for a given agent had to be vindicated by that agent's evaluative standpoint. Imagine, for instance, that Melanie is urgently driving her daughter to the hospital. Her daughter has been head-butted in the stomach by an angry ram while walking through a farmer's field and is suffering from internal bleeding. The child is moaning and wailing in the car; Melanie is driving like a maniac to reach the emergency room of the nearest hospital as soon as possible.

We might ask, 'What is to be said for Melanie performing this action?' The answer seems to be that it is good for the child: her child's life is at stake. And yet, under constructivism, if Melanie does not value her child's well-being and flourishing *at all*, then the fact that her child's life is at stake would not count as a reason for her to perform this action; it would not be a consideration weighing in favour of Melanie's action of driving her to the hospital. But, intuitively, we want to say that *all* agents have reasons – not necessarily overriding ones – to act so as to promote the well-being and flourishing of their offspring (assuming that they have offspring).

Intuitively, we want to say that any parent in Melanie's position, any parent whose child's life hangs in the balance, has reason to do whatever they can to save their child's life. The fact that Melanie's daughter is in the grips of a medical emergency is a normative reason for Melanie to get her medical attention as soon as possible, and this is true *irrespective* of what evaluations Melanie actually holds – just as it would be true for any parent in this situation, irrespective of what their actual evaluations happen to be.

According to Dorsey (2018: 578), what examples like this suggest is 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'. Dorsey also gives some other examples,

including 'the prevention of harm, the care of the sick, beneficence rather than maleficence, and so forth' (ibid.). It is important to note that although Dorsey is committed to the view that certain reasons are universal, he does not claim that such reasons are *overriding*. He makes the more modest claim that they are *significant* but not necessarily overriding (ibid.: 579). We can contrast 'significant' with 'trivial' here: if an agent has a reason of very little weight to help the poor merely for the sake of novelty, variety, or impressing their friends, say, then this is not the sort of reason that Dorsey has in mind.

The key takeaway here is that constructivism is *not* able to accommodate this intuition. Under constructivism, a particular fact cannot play a normative role unless that role is sanctioned in some way or other by the actual evaluative attitudes of the agent in question. Now, I think that Dorsey puts his finger on one facet of the radical contingency objection: under constructivism, normative practical reasons ultimately depend on the actual evaluative attitudes of individual agents, and because evaluative attitudes can vary dramatically across individuals, so, too, can the reasons that individual agents possess; this seems to fly in the face of our intuition that certain reasons – e.g. reasons not to creep up behind someone and shove them down a flight of stairs for a laugh – apply to *all* agents.

However, there is another facet to the radical contingency objection that Dorsey does not emphasise but which is worth emphasising. Under constructivism, it is not *just* that an agent's reasons depend, ultimately, on their actual evaluations (in concert with the relevant non-normative facts). Rather, it is that an agent's reasons depend, ultimately, on their actual evaluations at a *particular time*. So, the problem is not merely that no facts necessarily count as reasons for all agents; the problem is also that any facts that do count as reasons for action for a particular agent can, at any moment, *cease* counting as reasons due to a shift in that agent's evaluations. Under constructivism, reasons are like houses built on sand: if the sand starts to shift, they can come crashing down. In principle, an agent can go from having some moral reasons on Monday to having no moral reasons on Tuesday to having some moral reasons again on Wednesday.

Dorsey focuses on one aspect of the radical contingency objection, which is that under constructivism, there is no guarantee that different individuals will share the same core moral reasons. But there is a second aspect to the objection, which is that

under constructivism, there is no guarantee that the *same individual* will possess the same core moral reasons *across time* – and this also seems counter-intuitive. Both of these aspects are entailments of the fundamental dependency of reasons on attitudes that is posited by constructivism. At bottom, the concern here is that constructivism fails to capture the *permanence*, *stability*, *fixedness*, etc., that we intuitively think our moral reasons should have. One aspect of this concern relates to the fixedness of our reasons across individuals; the other relates to the fixedness of our reasons across time. They are two sides of the same coin.

To return to my earlier example, the fact that Melanie's child is suffering from internal bleeding after being head-butted in the belly by an irate ram counts in favour of getting that child medical attention as soon as possible. Intuitively, we want to *fix* this reason in place such that it is a reason for all agents who find themselves in this situation, no matter their evaluative starting points. We want this reason to 'stay put', as it were. Furthermore, if this reason *is*, in fact, a reason for an agent, we don't want it to suddenly cease being a reason due to a sudden and violent change in that agent's subjective motivational set. To illustrate this latter point, let me offer an example from the work of the Norwegian playwright Henrik Ibsen. The example is taken from Ibsen's classic play *A Doll's House*. The play centres around five main characters: Torvald Helmer, who is a bank lawyer; Nora, who is his wife; and their three young children.

Nora lives a suffocating and stifled existence as a housewife. Nora is much younger than her husband, Torvald, and he exercises a kind of 'benign' paternalistic control over her life (e.g. paying her a weekly allowance, telling her off as though she were a naughty schoolchild when she does something wrong, etc.). Throughout the play, Nora chafes under her husband's patronising manner. Her growing resentment towards him eventually culminates in a revelatory experience, wherein she resolves to cut ties with her husband completely and abandon her children so that she can start a new, free life for herself. Below is an extract from the end of the play; it is worth noting that the play caused an enormous scandal when it was first performed back in the 19th century (Ibsen 1981: vii–xiv):

NORA. [T]hat was the moment I realised that for eight years I'd been living with a stranger, and had borne him three children … As I am now, I am no wife for you.

HELMER. I still have it in me to change.

NORA. Perhaps … if you have your doll taken away.

HELMER. And be separated from you! No, no, Nora, the very thought of it is inconceivable.

NORA. [*goes into the room, right*]. All the more reason why it must be done.
[*She comes back with her outdoor things and a small travelling bag which she puts on the chair beside the table.*]

HELMER. Nora, Nora, not now! Wait till the morning.

NORA. [*putting on her coat*]. I can't spend the night in a strange man's room.

HELMER. Couldn't we go on living here like brother and sister …?

NORA. [*tying on her hat*]. You know very well that wouldn't last. [*She draws the shawl round her.*] Goodbye, Torvald. I don't want to see the children … As I am now, I can never be anything to them … Look, here's your ring back. Give me mine.

HELMER. That too?

NORA. That too.

HELMER. There it is.

NORA. Well, that's the end of that. I'll put the keys down here. The maids know where everything is in the house—better than I do, in fact. Kristine will come in the morning after I've left to pack up the few things I brought with me from home. I want them sent on …

HELMER. May I write to you, Nora?

NORA. No, never. I won't let you.

HELMER. But surely I can send you …

NORA. Nothing, nothing.

HELMER. Can't I help you if ever you need it?

NORA. I said 'no'. I don't accept things from strangers …
[*She goes out through the hall door.*]

HELMER. [*sinks down on a chair near the door and covers his face with his hands*]. Nora! Nora! [*He rises and looks around.*] Empty! She's gone! …

[*The heavy sound of a door being slammed is heard from below.*] (ibid.: 85–86).

Nora has undergone a radical change in her evaluations. She used to value being a faithful wife to her husband and a conscientious mother to her children, but now she does not value these things – or, at least, she values them much less than she did previously. New evaluations have sprung into being for her. Now, she deeply values living a free, independent, and self-actualised existence.

Under constructivism, these seismic changes in Nora's subjective motivational set would very likely bring about commensurate changes in her reasons. I say 'very likely' rather than 'definitely' because, as covered in Chapter 1, constructivism does not claim that an agent's reasons are determined *solely* by his or her evaluations. Rather, *X* is a reason for *A* to *Y* iff *X* is logically and instrumentally entailed from within *A*'s evaluative standpoint at a particular moment in time, the evaluative standpoint being composed of the full set of evaluations held by *A* – in either an occurrent or dispositional form – together with any non-normative facts that have a bearing on those evaluations. So, even if one's evaluations change radically, it could be that the facts of one's situation are such that one's reasons only undergo a comparatively small change.

But, for the sake of simplicity, let us put this caveat to one side and return to our example. Nora used to have reason, say, to be a dutiful, obedient wife to her husband. Now she does not. More controversially, she used to have reason to act in such a way as to promote the well-being and flourishing of her three young children. Now she does not – at least, this reason has been radically demoted in the hierarchy of her normative practical reasons. Nora used to place a high valuation on the welfare of her children. Now she regards their existence as parasitic on her own. Like her husband, her children are burdens imposed on her by the stifling conventions of 19th-century society, burdens that are inhibiting her *own* self-actualisation and flourishing.

The point of this example is to illustrate the second aspect of the radical contingency objection: that which relates to the permanence, stability, fixedness, etc., of our reasons *across time*. The thought here is that it seems counter-intuitive to say that Nora's reasons to nurture and care for her offspring can cease to exist if her

evaluations change to a sufficient degree (and then reappear if her evaluations swing back like a pendulum to their original orientation). Reasons, *moral* reasons especially, are not like quantum particles that pop in and out of existence depending on the composition of an agent's subjective motivational set; there is a stability, fixedness, and permanence to them – or so we intuitively feel.

Intuitively, we want to say that Nora's reasons in respect of her children persist over time and in spite of any turbulence that occurs in her motivational set. And we want to say this for any agent who finds themselves in a similar situation to Nora. The love of a parent towards their child is often held up as the quintessential example of unconditional love, love that persists in spite of the vicissitudes of life (Cordner 2016: 2).

This notion of unconditional parental love captures the inescapable nature of the moral reasons that apply to any parent with regard to their own children. A parent cannot escape these moral requirements in virtue of undergoing a sudden and violent attitudinal shift of the kind that Nora experiences. And yet, under constructivism, a parent *can* escape said requirements in just this way. Constructivism allows that even our core moral reasons are mutable across time, for they are anchored to our actual, mutable states of valuing. And this seems counter-intuitive.

3.7 CONSTRUCTIVISM STRIKES BACK: A RESPONSE TO THE RADICAL CONTINGENCY OBJECTION

The first thing to say in response to the radical contingency objection is that Dorsey actually misstates the constructivist position. Dorsey claims that, under constructivism, there are *no* reasons that are generally shared (2018: 577). This is not what constructivism claims. Constructivism claims that there are no reasons that are *necessarily* shared: all reasons are contingent in the sense that they are anchored to the contingently held, actual evaluative attitudes of valuing creatures.

The fact that there are no reasons that are *necessarily* shared by all human beings does not entail that there are no reasons that are *generally* shared or even *universally* shared. Reasons could be generally or universally shared for purely contingent reasons, such as deep and widespread similarities in the evaluative standpoints of agents. Indeed, given our shared evolutionary history here on planet

Earth and the ways in which evolutionary pressures have shaped our needs and wants, we can expect there to be widespread – albeit not perfect – convergence in our normative practical reasons, including our moral reasons. So, contrary to what Dorsey (2018: 577) alleges, we can *expect* that, under a constructivist view of things, there will be some reasons that are generally shared by human beings.

To see this, let us start with the fact that we are social primates who evolved in a particular social context: that of hunter-gatherer bands numbering no more than 150 individuals (Harari 2011: 29–30). Obviously, most of us in the modern world no longer live this way, but that is a relatively recent phenomenon. From 2.5 million years ago, when humans first evolved, to 12,000 years ago, when the Agricultural Revolution took place, we lived in roaming bands of hunter-gatherers (ibid.: 6, 89–90).[22]

Those millions of years spent evolving in the context of small hunter-gatherer communities had a profound impact on our moral psychology. Hunter-gatherer bands were tight-knit and intimate: everyone knew everyone else, and there was a high degree of mutual dependence (Pinker 2022: 68). In such a social environment, pro-social, cooperative, other-regarding, and even altruistic behaviour almost always redounded to the benefit of the individual who exhibited such behaviour. Anti-social, exploitative, and selfish behaviour, on the other hand, would quickly earn one a reputation as a bad and untrustworthy person, and this would, in turn, lead to shunning and ostracism by other members of the tribe (Krebs 2008: 154ff.).

For example, if one member of a hunter-gatherer tribe accepted a share of the food that other members of the tribe had hunted or gathered – e.g. meat from a successful hunt or berries from a successful harvest – but declined to share any of the food that he himself had hunted or gathered, then the next time he came asking other members of the tribe for some food, they would remember his lack of reciprocity and would likely refuse him. And if this member of the tribe were to continue behaving in a selfish, non-reciprocal way, then he might gradually come to

---

[22] Interestingly, Steckel (2008: 148–149) has shown that compared to hunter-gatherers, the first city dwellers suffered from much higher rates of anaemia, infection, and tooth decay and, on average, were almost two and a half inches shorter in height. Some biblical scholars – e.g. Kugel (2007) – think that the story of the Fall from the Garden of Eden was a cultural memory of the transition from a hunter-gatherer lifestyle to an early agricultural one.

be treated as a social pariah or even be exiled from the tribe altogether. He would have to become self-sufficient in terms of hunting and gathering food.

It is unlikely that, in the unforgiving Palaeolithic environment of our ancestors, an individual would have been able to maintain this kind of self-sufficiency for very long: not every hunt is successful, especially when one is hunting alone; not every effort at berry-picking will yield fruit; and there are only so many meals that one can miss before one starves to death. As Charles Darwin (1874: 102) puts it, 'the individuals which took the greatest pleasure in society would best escape various dangers; whilst those that cared least for their comrades, and lived solitary, would perish in greater numbers'. If an individual dies as a consequence of social ostracism, then, presuming that he has not spawned any offspring, his selfish genes will die with him.

For millions of years, evolution selected for humans who were able to cooperate effectively with others in their tribe. One of the key forms of cooperation that was selected for was *mutualism*, which is when two or more agents pursue a goal that they are unable to achieve on their own or a goal that they can achieve more effectively by working together (Krebs 2008: 154–155). Examples of mutualism include raising offspring, grooming, and hunting large game, such as woolly mammoths.

Another key form of cooperation was *long-term social investments* (Krebs 2008: 154–155). This is when members of a group form mutually beneficial friendships with those who possess complementary abilities or resources. For example, someone who was very strong but not very bright might make friends with someone who was very bright but not very strong: a partnership of brawn and brains, the meathead athlete and the sickly scholar. The athlete could offer the scholar protection against aggressors and bullies, and in return, the scholar could use his superior intelligence to help the athlete solve difficult problems, such as how to start a fire when all the wood was still damp from the recent rainfall.

A member of a tribe is unable to interact with everyone else in the tribe to the same degree, so he must try to fill his association niche by making friends with those who are most willing and able to foster his fitness. It was in our Palaeolithic ancestors' long-term interest to 'suffer the relatively small costs of helping friends over long periods of time if such acts increased the likelihood that their friends …

would be there for them when their fitness was in jeopardy [e.g. when they fell sick and needed someone to take care of them]' (Krebs 2008: 154–155).

In addition to natural selection for those genes which disposed humans to cooperate effectively with others in their tribe, there was also sexual selection and kin selection for those genes which disposed humans to behave *altruistically* in certain contexts. What I mean by altruistic behaviour is behaviour that is 'costly to the actor and beneficial to the recipient' in terms of 'lifetime personal fitness consequences' (Kurzban, Burton-Chellew, and West 2015: 578). The term 'personal fitness' refers to the success of an individual in transmitting copies of their genes to future generations through reproduction (Rodrigues and Gardner 2021: 347). When an individual behaves altruistically, they pay a fitness cost (i.e. it reduces the likelihood, by however small an amount, that they will survive and produce offspring), while as the recipient enjoys a fitness benefit (i.e. it increases the likelihood that they will survive and produce offspring). There are many examples of altruistic behaviour in the animal kingdom, such as when an individual bee sacrifices its life for the sake of the hive of which it is a member (Ananth 2005: 219).

At first glance, altruistic behaviour seems to go against the 'survival of the fittest' ethos of Darwinian natural selection, so how might it have evolved in humans? There are two main ways: sexual selection and kin selection. In sexual selection, traits that are 'attractive to the opposite sex, and that therefore increase the probability of mating, will evolve even if they are not optimal for survival' (Krebs 2008: 156). The classic example of this is the flamboyantly colourful tails of male peacocks. These tails make male peacocks far more visible than they otherwise would be and, therefore, more vulnerable to predation, but they are highly attractive to female peacocks.

When it comes to *human* sexual selection, members of both sexes rank qualities related to altruism near the top of their lists of attractive qualities in the opposite sex (Buss 2004; Miller 2007). Potential mates prefer partners who possess costly altruistic traits for two main reasons. First, the traits signal that the individual in question will treat any potential mates and progeny altruistically, which is a good thing from their point of view as the beneficiaries of that altruism. Second, altruistic traits indicate that the individual who holds them has good, healthy genes that render them strong enough to survive *in spite of* the high costs of behaving altruistically. In

other words, altruistic traits are a way of signalling to others that one is so overflowing with genetic fitness that one can *afford* to help others at one's own expense.[23]

In addition to sexual selection for altruistic traits, there was also kin selection for these traits. The basic idea behind kin selection is that genes may increase in frequency by 'guiding the development of mechanisms that dispose individuals to sacrifice their bodies (the vehicles transporting the genes) in order to preserve other bodies that share copies of their genes by descent' (Krebs 2008: 156). This construal of kin selection is implicitly predicated on the idea that genes are the fundamental, elemental units of natural selection (Dawkins 2006). There is abundant evidence that members of many species, including *Homo sapiens*, are disposed to sacrifice their personal and biological interests for the sake of their children, siblings, parents, etc. (Alcock 2005; Kurland and Gaulin 2005).

The extent to which one is willing to make altruistic self-sacrifices for kin is positively correlated with the degree of relatedness. For instance, Segal, Hershberger, and Arad (2003) found that identical human twins were more likely to behave altruistically towards one another than fraternal twins were. I should say that although I have focused on the evolution of traits relating to altruism, a very similar story could be told about sexual and kin selection for other traits that are commonly perceived to be virtuous, such as fairness, forgiveness, fidelity, etc. Fairness and fidelity, for instance, are costly to maintain in terms of fitness but are attractive to potential mates, while as perceived moral vices, such as capriciousness or sexual profligacy, are less costly to maintain but are off-putting to potential mates.

The key takeaway from this foray into evolutionary psychology is as follows. For most of our evolutionary history there have been strong selection pressures

---

[23] Given that altruistic traits have been sexually selected for, it might seem strange to describe them as fitness *costs*. Recall that 'fitness' refers to the success of an individual in transmitting copies of their genes to future generations through reproduction (Rodrigues and Gardner 2021: 347). Altruistic traits do not, *all things considered*, reduce an individual's fitness: they are not a *net* negative. Rather, they reduce one's fitness in the short run or in specific ways. For example, if an individual makes a habit of giving away food to hungry members of his tribe, then this will mean that he has less food with which to nourish himself. However, this generous and self-sacrificing behaviour on his part may help him to attract a desirable mate, which, in the long run, will enhance his fitness.

favouring the development of pro-social, cooperative, altruistic, and other-regarding traits – at least when it comes to members of one's own tribe. An important point concerning these traits is that one must be able to convince others that one *genuinely does* possess them. If one is wooing a potential mate and that mate is looking for evidence of altruistic traits in oneself, then one must be able to convince that person that one genuinely does have those traits and is not merely *pretending* to have them.

Studies have found that humans are very astute when it comes to identifying who really possesses valued traits, such as kindness, and who does not (Krebs 2008: 155). From an evolutionary perspective, the 'easiest' way – i.e. the way that is least costly in terms of personal fitness – for an individual to convince others that he is pro-social, cooperative, altruistic, other-regarding, etc., is for him to *genuinely be* pro-social, cooperative, altruistic, other-regarding, etc. (Sobel 2016: 24, 30).

The alternative is for that individual to engage in a highly elaborate and ongoing deception, wherein he behaves in such a way as to convince others that he is pro-social, altruistic, etc., but, in reality, does not possess these traits. Given that we are quite good at detecting frauds in this area, the deception really would have to be very elaborate to be consistently convincing to others. It simply would absorb too much of one's cognitive bandwidth to engage in an ongoing pantomime of this kind; it is simpler and more efficient for the traits to be genuine.

The upshot of this is that the vast majority of actual human beings are genuinely pro-social, cooperative, altruistic, and other-regarding to some degree. There is some empirical evidence in support of this claim. I hasten to add, however, that this empirical evidence is of an indirect rather than direct sort: it tends to focus on people's *judgments* about what their other-regarding, etc., duties are, as opposed to focusing on their other-regarding *behaviour* itself. (I am interpreting people's judgments about what moral duties they have as judgments about what moral *reasons* they have.) Studies have shown that even across widely different cultures, almost all people judge that they have a duty to: (a) fulfil their social obligations (Berkowitz 1972); (b) reciprocate with others (Gouldner 1960); (c) refrain from hurting others gratuitously (Turiel 2002). Studies also show that almost everyone endorses some version of the Golden Rule (Hauser 2006; Wright 1994).

To be clear, I am not claiming that evolution has equipped us with other-regarding motives that are sufficiently strong to result in action in every case or even in most cases. There are many contexts in which such motives might be outweighed by more self-serving ones. The voices of the better angels of our nature are often drowned out by the hooting of our lesser demons. However, these other-regarding motives are significant: they are strong enough to impel action at least some of the time. If they were not, if they were merely trivial, then they would not have been selected for; such motives were (naturally) selected for because they often result in behaviour that redounds to the all-things-considered fitness of the organism that possesses them. As Blackburn (1993: 168) puts it, 'The evolutionary success that attends some [conative states] and not others is a matter of the *behaviour* [emphasis added] to which they lead … it is the direct consequences of the pressure on action that matter. Evolutionary success may attend the animal that helps those that have helped it, but it would not attend any allegedly possible animal that thinks it ought to help but does not. In the competition for survival, it is what the animal does that matters.'

One important caveat here is that we evolved to feel other-regarding motives primarily towards our kin, friends, potential mates, and fellow tribesmen. As I discussed earlier, one's own tribe almost always consisted of fewer than 150 individuals, all of whom one would have known and interacted with on a regular basis. Other-regarding motives are much weaker in the case of strangers, especially those strangers whom one is never likely to have a face-to-face encounter with, such as those living on a different continent from oneself and who speak a different language and have different customs (Slovic 2007).

But even with this caveat on board, it is still true to say that '[m]ost humans, even most murderers, will have some significant sympathy for ordinary fellow humans' (Sobel 2016: 23–24). As Sobel (ibid.) puts it, 'Almost everyone we know cares not only to avoid inflicting misery on others but also to have their behavior be justifiable to others … The truth is very few of us really want to harm many others and almost all of us, as Hume told us, resonate in sympathy with the vast bulk of humanity [though this sympathy tends to become diluted the further out we go from our immediate social network]'.

Let us bring this back to the radical contingency objection against constructivism. Recall that Dorsey's (2018: 577–578) gripe with constructivism is that the view supposedly 'denies that there are any reasons that are generally shared', and this seems contrary to our intuition that there are certain reasons that 'we all share'. What makes this point a bit confusing is that, as discussed earlier, Dorsey (ibid.) seems to blur the distinction between reasons that are generally shared and reasons that are *universally* shared.

As I hope to have shown over the preceding paragraphs, Dorsey is wrong to think that under constructivism, there are no reasons that are *generally* shared among all actual human agents. Our contingent evolutionary history as a species has sculpted our moral psychology such that the vast majority of us have significant motivation to act in ways that are pro-social, cooperative, altruistic, other-regarding, etc. These motivational pressures will naturally be reflected in our evaluative attitudes, which are just conative states: states of taking certain things/options in the world to be good or bad, right or wrong, better or worse, worthy or worthless, required or optional, virtuous or sinful, honourable or dishonourable, valuable or disvaluable, etc. Under constructivism, truths about what reasons an agent possesses are determined, at least in part, by whatever evaluations that agent holds. Given that there is widespread convergence in the evaluative standpoints of human beings, we can expect that, depending on the relevant non-normative facts, there will be widespread – though probably not *universal* – convergence in what reasons human beings possess.

To make this more concrete, let us focus on Dorsey's (2018: 578) example of a reason that we supposedly all share: 'the fact that children need minimal care to survive and flourish is a reason for anyone to provide such care for their (or indeed anyone's) children'. The constructivist can tell a convincing story, backed up by plenteous evidence from evolutionary biology, about why, as a matter of contingent fact, almost all human parents have strong reasons to act in such a way as to promote the well-being and flourishing of their offspring.

Earlier, I discussed kin selection and how genes may increase in frequency by 'guiding the development of mechanisms that dispose individuals to sacrifice their bodies (the vehicles transporting the genes) in order to preserve other bodies that share copies of their genes by descent' (Krebs 2008: 156). The level of altruistic self-

sacrifice that one is prepared to engage in varies with the degree of relatedness, in genetic terms, to the person for whom the sacrifices are made (Segal, Hershberger, and Arad 2003). Unsurprisingly, parents are prepared to make enormous personal sacrifices for the sake of their children, who contain copies of half of their genes, and who will ensure that those genes are propagated to future generations.

Of course, parents are not *motivated* to make these sacrifices by considerations relating to genetic propagation; genes pull the strings of their vessels via conative states, such as parental feelings of love and sympathy for one's offspring (Dawkins 2006; Slovic 2007). Such feelings are extremely powerful and deeply rooted. Indeed, Batson et al. (2005) suggest that maternal care, in particular, was the evolutionary precursor of all other forms of human sympathy. With this in mind, we can expect virtually all actual human parents to hold evaluations in respect of their offspring that (indirectly and in concert with the relevant non-normative facts) provide them with significant reasons to act in such a way as to promote the well-being and flourishing of their offspring – e.g. by making sure they are well-fed and not letting them wither away from malnutrition.

We can expect that, under a constructivist view of things, such reasons will be very, very widely shared (among human parents). However, constructivism cannot guarantee, in advance, that such reasons will be *universally* shared; they may be, but if they are, then that will come down to contingent similarities in the evaluative standpoints of human beings. There might be parents out there – even if only a few – who are so morbidly selfish that they do not care at all about the welfare of their children; depending on what the relevant non-normative facts are, it might not follow from such parents' evaluative standpoints that they have moral reasons to care for and nurture their children – at least, the constructivist cannot *guarantee* that they will have these reasons. Having said this, even if these selfish parents lacked *moral* reasons to care for their offspring, they would very likely have *non-moral* normative reasons to do so (e.g. because neglecting their children would damage their reputation and standing in their community and could lead to legal repercussions).

If our intuition is that all human agents, *without exception*, have a reason, say, to provide at least a minimal level of care 'for their (or indeed anyone's) children', then

117

constructivism cannot accommodate this intuition (Dorsey 2018: 578).[24] And the same goes for any other reasons that we intuitively feel to be exceptionless in this way. In this regard, constructivism is revisionist: it revises one of our pre-theoretical moral intuitions. However, I hope to have shown in this section how the revisionist element of the theory is tolerably rather than intolerably so. We have good grounds for thinking, based on various contingent but deeply-rooted features of our evolved moral psychology, that there will be widespread convergence in the reasons that actual agents possess. So, the constructivist might not be able to secure perfect universality in our core moral reasons, but he can at least secure a very widespread generality in those reasons.

3.8 ON THE PERMANENCE OF OUR MORAL REASONS

The preceding section evaluated one half of the radical contingency objection: it is the half that Dorsey (2018: 577–578) focuses on, and it concerns the fixedness of our core moral reasons *across individuals*. Let us now consider the second half of the objection, which I sketched out in §3.6.

Recall that this second half of the objection concerns the fixedness of our core moral reasons *across time*. To recap, the problem is that under constructivism, any facts that *do* count as normative practical reasons for a particular agent can, at any moment, cease counting as reasons due to a shift in that agent's evaluative attitudes. Under constructivism, reasons are like houses built on sand: if the sand starts to shift, they can come crashing down. In principle, an agent can go from having some moral reasons on Monday to having no moral reasons on Tuesday to having some moral reasons again on Wednesday. The concern here is that the theory fails to capture the permanence, stability, fixedness, etc., that we intuitively feel our moral reasons should have *across time*.

Back in §3.6 I quoted an extract from Ibsen's classic play *A Doll's House*, wherein Nora, an oppressed housewife, resolves to cut ties with her husband, Torvald, and their three young children. Nora makes this resolution after undergoing a transformative shift in her evaluations. This example was meant to illustrate the point

---

[24] For what it is worth, this is not an intuition that I personally share, but let us assume for the sake of argument that Dorsey is correct in claiming that this *is*, in fact, one of our strongest pre-theoretical intuitions in this area.

that, under constructivism, an agent's reasons can oscillate in tandem with oscillations in their evaluations.

Nora used to have a reason, say, to act in such a way as to promote the well-being and flourishing of her three young children. But after the volcanic eruption that took place in her subjective motivational set, it is no longer entailed from within her evaluative standpoint that she has such a reason (let us just assume for the sake of argument that this is true). But if one week after abandoning her husband and children Nora suddenly has a change of heart – i.e. if her states of valuing swing like a pendulum back to their original orientation – then she will, once again, have a reason to act in such a way as to promote the well-being and flourishing of her children. This sort of mutability in our reasons across time seems disconcertingly revisionist with respect to our pre-theoretical intuitions in this area – or so the objection goes.

What does the constructivist have to say in response to this? The first thing to note is that under constructivism, our reasons are not quite as prone to wild fluctuations as they might seem. The conative states that go towards fixing ethical truth, under constructivism, are attitudes of *valuing*, as opposed to more ephemeral mental states like whims or fancies.

Recall from Chapter 1 that an attitude of valuing (hereafter: evaluation) is a conative state of taking certain things/options in the world to be good or bad, right or wrong, better or worse, worthy or worthless, required or optional, virtuous or sinful, honourable or dishonourable, valuable or disvaluable, etc. It is the state of mind of judging, at least implicitly, that certain things/options 'call for, demand, or provide reasons for others' (Street 2010: 366). Such judgments need not be held in an explicit or occurrent form: they could be implicit or dispositional. An example of an evaluation under this construal would be the following judgment made by Blanche DuBois in *A Streetcar Named Desire*: 'Deliberate cruelty is not forgivable. It is the one unforgivable thing in my opinion' (Williams 2009: 93).

We can distinguish this state of mind – that of evaluating certain things/options in the world – from the state of mind of *mere* desiring (which would encompass states like passing whims or fancies). The state of *mere* desiring involves a feeling of being pleasantly attracted towards some object/state of affairs (Street 2012: 43). For

119

example, if one is walking down the street and happens to chance upon a charming little artisanal bakery, one might desire a slice of a chocolate torte that is on display in the window.

Under this narrow, familiar sense of the term 'desire', the notions of *desiring* and *valuing* can come apart: an agent can desire something without valuing it (Sobel 2016: 4). To stick with the chocolate torte example, an agent could desire the sumptuous cake without evaluating, say, that it would be good or worthy to eat it. If the agent in question suffers from high cholesterol and obesity and is trying to get his health back on track, then his feeling of pleasant attraction towards the cake might be quashed by his grim determination to stick to a healthy diet.

The key point here is that mere desires do not, under a constructivist view of things, go towards determining truths about what moral reasons one possesses; it is only *evaluations* that play this role. Mere desires are transient, shallow, and undisciplined in a way that renders them unsuitable for use in the constructivist method for generating ethical truth. Such states do not capture what we *really* want out of life; they are fancies that occasionally flutter through the mind like butterflies, capturing our attention for a short while before passing away.

When I say that mere desires are 'undisciplined', what I mean is that they do not conform to the instrumental requirements that govern practical reasoning: as a constitutive matter, they do not aim at means-end coherence. So, one can desire some end without desiring what one is fully aware is the necessary means to that end. One can desire a luxury beach holiday in the Caribbean, say, without desiring to work overtime for the next six months in order to save up sufficient funds to pay for that holiday. As discussed in §1.1, this is *not* something that one can do when it comes to evaluations, which do, as a constitutive matter, aim at means-end coherence.

The language of desire, understood in this ordinary, common-parlance sense, is inadequate to capture the kind of emotional and phenomenological complexity involved in those states of mind that confer normativity on the world. States of *valuing*, by contrast, have a depth and stability to them that states of mere desiring do not. As Street (2012: 44) puts it, the state of mind of valuing is characterised by all of the 'range, nuance, and depth of human emotion and feeling'. A person's values

are reflected in such actions as 'getting up at 4 a.m. to finish grading papers; breaking bad news to a friend; or risking death to fight an authoritarian regime' (ibid.).

For example, the values that many of the citizens of Ukraine hold are reflected in the actions that they are taking to defend their country against Russian imperium, actions which often involve risking life and limb. The people of Ukraine – and others around the world who find themselves in a similar situation, resisting tyranny – have affective, sometimes intensely emotional, experiences of such actions as being *demanded* or *called for*. But these actions do not call out to them in a way that is pleasant, as the term 'desire' connotes, but in a manner that involves 'something more like anxiety or sickness at the thought of not doing them, feelings of gritty determination to preserve what one cares about most in spite of everything' (Street 2012: 44). Unlike mere desires, then, attitudes of valuing can be directed at objects or states of affairs that one finds 'unpleasant, terrible, anguish-inspiring, and so forth' (ibid.).[25]

The key takeaway here is that under constructivism, the states of mind that go towards determining truths about one's reasons are deeply rooted and central to who we are as people; they are not whimsical or unserious states. Let me offer an example to illustrate this point. Imagine that one is driving one's car along a road, minding one's business, when all of a sudden one feels a strong urge to pull an ugly face at other drivers along the road, and even at pedestrians that one passes by. In life, giddy whimsies of this kind sometimes overtake us, and sometimes they are strong enough to result in action. But conative states of this sort do not, under a constructivist view of things, go towards fixing ethical truth.

Just because one is seized by a sudden desire to pull ugly faces at people does not mean that one's reasons will automatically shift to accommodate this new and

---

[25] The term 'desire' is often used in a far more expansive sense than that which I have been focusing on here: it is sometimes used as an 'umbrella term' to refer to a whole class of favouring attitudes, such as 'liking … preferring, wanting, loving, caring for, being devoted to, cherishing, craving, etc.' (Sobel 2016: 3–4). We might think of these as *broad desires*, as opposed to *narrow desires*, which involve a feeling of being pleasantly attracted towards some object/state of affairs. *Broad* desires could qualify as evaluations, providing that they contain, at least implicitly, some judgment about whether certain things/options in the world are good or bad, right or wrong, better or worse, worthy or worthless, required or optional, virtuous or sinful, honourable or dishonourable, valuable or disvaluable, etc. So, evaluations can overlap with broad desires, but they do not overlap with what I have been calling 'mere desires'.

mysterious state of mind. Generally speaking, one's *evaluations* have much greater stability than one's (mere) desires. And this fact translates, other things being equal, into greater stability in one's *reasons*. So, to revise my earlier analogy, reasons are not, under constructivism, as unstable and precarious as houses built on sand. Rather, they are more like houses built near a riverbank. That is, they are generally solid and stable, but if the river floods – i.e. if one's evaluations undergo some extraordinary transformation – they can be swept away by the raging torrents.

In truth, even this riverbank analogy overstates the changeability of one's moral reasons across time. Under constructivism, it is not as though each moral reason that one has is attached, as if by a piece of string, to individual evaluations in one's subjective motivational set, such that if a single evaluation changes, it yanks out a reason along with it. It is not *individual* evaluations that matter under constructivism. Rather, it is the *full set* of one's actual evaluations that are determinative, in concert with the relevant non-normative facts, of truths about one's reasons.

We can imagine a situation in which a sub-set of an agent's full set of evaluations changes over time without this triggering any corresponding change in the reasons he possesses. Imagine that we have two cousins: Matsuoka Bungō and Okamura Heiroku. Let us say that Matsuoka and Okamura were very close growing up. They spent much time together and always got along well; they seemed more like brothers than cousins. But as they grew older, Matsuoka's personality began to change for the worse. Where before he was good-natured, kind, and patient, he became irascible, spiteful, and impulsive. By the time Matsuoka reached manhood, he had morphed into a genuinely unpleasant person, someone who led a debauched existence and who, through his reckless and insensitive behaviour, created much suffering for his family – e.g. by getting into drunken bar fights, gambling his money away, keeping company with lewd and saucy women, stealing from his family and friends, getting into constant run-ins with the law, etc.

We can imagine Okamura becoming increasingly horrified and alienated by the behaviour of his childhood friend. This would, presumably, be reflected in changes in his evaluations across time. At one point in time, say, he might have held (at least implicitly) the following evaluations. 'Matsuoka is my good friend, and I ought to do what I can to make him happy. We should always give our friends the benefit of the doubt and be willing to forgive them if they act in ways that we do not like. For

friendship to really mean something, it must endure even when times are hard. It would be wrong for me to abandon Matsuoka, even though he is heaping shame and misery on our family.'

Now, we can imagine Matsuoka doing something so outrageous and offensive – e.g. robbing someone at knifepoint on the street – that the above evaluations either drop out of Okamura's motivational set entirely or are severely diminished in strength. That is to say, there may be a 'final straw' moment where Okamura finally decides that Matsuoka has behaved too despicably to be forgiven.

Despite this, it would not be surprising if, under constructivism, there was *still no change* in Okamura's (moral) reasons with respect to his cousin. These reasons might be held in place, as it were, by evaluations in Okamura's motivational set that have *not* changed over time. For instance, even if Okamura no longer evaluates that 'Matsuoka is my good friend, and I ought to do what I can to make him happy', he may still hold other evaluations that have a bearing on his reasons in this domain, such as, 'It is right and worthy to stand by one's extended family, no matter how bad things get', or, 'One must preserve one's family honour even if it means making considerable personal sacrifices.'

The point is that the balance of the full set of one's evaluations will not necessarily be disrupted by changes in individual evaluations or clusters of evaluations across time. By analogy, if one starts removing individual bricks from a watchtower, it may be a while before the tower as a whole collapses. Okamura may continue to have a reason – though not necessarily an overriding one – to help Matsuoka in whatever way he can to rehabilitate himself and to get his life back on track (e.g. by offering emotional or financial support).

Furthermore, under constructivism, it is not *just* one's evaluations that go towards fixing ethical truth. The relevant non-normative facts have a crucial role to play as well. To see this, let us return to the example of Matsuoka Bungō and Okamura Heiroku. Imagine that Okamura begins to suspect that there is some medical cause underlying his cousin's downward spiral. Okamura thinks, *Maybe my cousin, Matsuoka, is suffering from some kind of neurological condition that is driving his self-destructive behaviour? Yes, perhaps he has brain damage – that would explain a lot.*

After much cajoling, Okamura convinces Matsuoka to visit a neurologist. The neurologist runs some diagnostic scans and discovers that Matsuoka has a tumour in his brain the size of a peach. The tumour is located in the medial prefrontal cortex, a region of the brain responsible for the control of emotion and behavioural impulses (Harris 2012: 142). Damage to the medial prefrontal cortex is associated with a 'variety of deficits including poor impulse control, emotional blunting, and the attenuation of social emotions like empathy, shame, embarrassment, and guilt … the ability to behave appropriately toward others tends to be disrupted' (ibid.: 124).

This single fact – that Matsuoka has a tumour in his brain – changes everything. Matsuoka no longer seems like a callous ruffian but like a victim of bad luck. He is a person who had the horrible misfortune of developing a brain tumour as an adolescent, a brain tumour which then went undiagnosed for years, all the while slowly warping his personality.

Under constructivism, facts of this sort, facts that are relevant to one's evaluations, are part of the raw materials that are used to construct truths about one's reasons. Okamura presumably thought, up until the trip to the neurologist, that his cousin's dissolute behaviour was a result of his having a genuinely bad character, an authentically knavish disposition. Many of his evaluations vis-à-vis Matsuoka would have been implicitly predicated on this theory of mind. But the discovery of the brain tumour is exculpatory and perspective-altering. Matsuoka now seems like a victim – or perhaps like some dangerous animal. We might compare him to a dog that is placid and friendly to begin with but which, unbeknownst to its owners, one day catches rabies and, thereafter, starts behaving in an increasingly deranged and aggressive manner. Once one discovers that the dog has rabies, one is no longer inclined to *blame* the dog for its behaviour in any deep sense. If anything, one feels sorry for it.

The incorporation of relevant non-normative facts into the constructivist method can, in many cases, increase the stability, permanence, fixedness, etc., of an agent's reasons across time. For instance, from the moment that Matsuoka first developed a brain tumour and his personality began to change, Okamura had a strong moral reason to get his friend the medical help he needed.[26] And this fact, this reason,

---

[26] And this is true regardless of what Okamura *thought* his reasons were during this time. As covered in §2.1, there are a variety of ways, under constructivism, in which an agent can be mistaken about what their reasons are.

remained stable across time. It remained stable right through Matsuoka's degeneration. It remained stable through all of the changes in the evaluations that Okamura held with regard to his cousin.

Of course, there are situations in which the relevant non-normative facts might have the exact opposite effect: they might suddenly *destabilise* one's reasons. For example, if, unbeknownst to oneself, one's boiler starts leaking poisonous carbon monoxide gas in the middle of the night, then one suddenly has most reason to immediately get oneself and one's family out of the house before inhalation of the gas causes death by suffocation. But I think most people would agree that in such cases we intuitively *want* our reasons to be able to shift in response to changes in the relevant non-normative facts, and it is a virtue of constructivism that it can accommodate such changes.

What I hope to have shown in this section is that constructivism is able to defang the objection that it is unable to capture the stability that we intuitively feel our core moral reasons should have across time. One need not worry that, under the view, one's reasons will be wildly fluctuating from moment to moment, popping in and out of existence like quantum particles. On the contrary, one can expect a reassuring and intuitive degree of continuity in one's reasons across time.

<p style="text-align:center">***</p>

In this chapter I evaluated three core objections to constructivism. This included an objection put forth by Russ Shafer-Landau (2003: 41–43), two variants of the hypothetical amoralist objection (Sobel 2016: 19–20), and finally, the radical contingency objection (which I split into two discrete halves) (Dorsey 2018: 579). I hope to have shown in this chapter that, in the case of all three of these objections, constructivism is either able to neutralise the objection entirely or at least to greatly diminish its force (such as to render the theory tolerably rather than intolerably revisionist). And this fact, when taken together with the advantages of the view that I laid out in Chapter 2, goes towards showing that constructivism is an attractive and tenable meta-ethical position.

To quote Shafer-Landau (2003: 43), 'Theory choice in metaethics, as elsewhere, is a matter of noting comparative strengths and weaknesses'. Every meta-ethical theory has its pros and cons. Every portrait hanging in the gallery of meta-ethical theories has its beauties and its blemishes, and one must balance these beauties

and blemishes in coming up with an overall assessment of any given portrait. For instance, constructivism may be somewhat revisionist when it comes to our intuitions about paradigmatic cases of moral and immoral behaviour, but, as covered in §2.2, it arguably does a better job of conserving the practical, action-guiding character of sincere moral judgments than some of its meta-ethical rivals.

When assessing the plausibility of constructivism, then, one must try to be like the statue of Lady Justice that stands outside of courthouses. One must stand with the scales in one's hands and carefully place the pros of the position on one side of the scales and the cons on the other side, and then see which tips the balance. In the next chapter, I will try to rustle up a few more pros to put on the scales.

4. Enlightened Constructivism: Do Peak Experiences and Meditation Have a Role to Play in the Construction of Ethical Truth?

I want to explore whether, by introducing an element of idealisation into constructivism, we can render the view a more attractive prospect to those philosophers who have an ambient level of scepticism about any form of ethical subjectivism – which happens to be a great many philosophers (Sobel 2016: 9). To be more specific, I want to explore whether we can draw on *peak experiences* to create a modest and defensible form of idealisation vis-à-vis the constructivist method (I will define peak experiences in detail in §4.1; for now, I just want to sketch out the contours of this chapter). The rough idea is to restrict the states of valuing that are fed into the method for constructing truths about one's moral reasons to those states of valuing that one holds, in either an occurrent or dispositional form, during peak experiences. I call this experimental view 'enlightened constructivism'.

One of the key motives for pursuing such an idealisation – besides that it is interesting – is to strengthen the force of one's moral reasons under constructivism. Recall that in §3.7 of the previous chapter I discussed how, due to various contingent features of our evolutionary history as a species, virtually all human beings have significant reasons to act morally (i.e. to act in a way that is pro-social, cooperative, altruistic, and other-regarding). I argued that this sort of contingent, evolved convergence in our moral reasons is of a sufficient degree to broadly satisfy our intuitive expectations of what extensional adequacy would look like in this domain.

But we can think of there being two planes to the extension that we intuitively expect our moral reasons to have: the horizontal and vertical. The horizontal plane relates to how *widely* certain reasons apply to actual or hypothetical agents. For example, do we intuitively think that reasons to care for and nurture one's offspring are universal in that they apply to all possible agents, or do we merely think that such reasons apply to virtually all actual human agents? I argued in §3.7 that constructivism comes very close to achieving extensional adequacy along this horizontal plane, and if it is revisionist to some degree, then it is tolerably rather than intolerably so.

The vertical plane of extensional adequacy, by contrast, relates to the strength or weight of our moral reasons. As discussed in §1.3, the strength of a normative practical reason for action co-varies, under constructivism, with the strength of the relevant evaluations in an agent's subjective motivational set. By 'relevant', I just mean whatever evaluations go towards determining truths about a particular reason that one possesses; these evaluations may be held in an occurrent or dispositional form. So, *other things being equal*, the stronger one's relevant evaluations, the stronger one's reasons. Crucially, 'other things' include the relevant non-normative facts, which often have a decisive role to play in determining truths about what reasons one possesses (see §3.8). Constructivism locates the moral within the domain of the normative, so an agent's moral reasons are a subset of the full set of their normative practical reasons more generally. Under constructivism, one's moral reasons are internal reasons in the sense that they are anchored to – but not directly determined by – one's subjective motivational set.

I argued in Chapter 3, §3.7, that the moral reasons that the vast majority of us possess, under constructivism, are not merely trivial. Rather, they are of significant weight. This seems in keeping with our intuitions. Intuitively, we do not expect moral reasons to be necessarily overriding (Dorsey 2018: 579). We want them to be *significant*, certainly, but they need not be overriding. This being said, one might worry that although constructivism can generate moral reasons that are of significant weight and are near-universally shared by all actual human agents, these reasons still do not have as great a weight as one would intuitively want them to – or, at least, as those who are of a more realist meta-ethical bent would intuitively want them to.

For example, although it is true that virtually all human agents have evolved to feel strong *motives* to act in a way that is other-regarding when it comes to members of their own social tribe – e.g. family, friends, neighbours, members of one's local community, work colleagues, potential mates, etc. – one might worry that because these motives are much weaker in the case of strangers, especially those strangers whom one is never likely to have a face-to-face encounter with, such as those living on a different continent from oneself and who speak a different language and have different customs, this would, under a constructivist view of things (and depending on what the relevant non-normative facts are), probably translate into our having weaker

moral *reasons* when it comes to acting towards strangers than acting towards family, friends, neighbours, etc.[27]

With this in mind, I want to explore whether introducing a peak-experience-based element of idealisation into the constructivist method is a plausible means of bolstering or 'shoring up' the strength of our core moral reasons. The rough idea is that peak experiences strengthen our evaluations that are more pro-social, cooperative, altruistic, and other-regarding in orientation, and if we draw on these 'peak' evaluations in the construction of ethical truth, then this would – depending on what the relevant non-normative facts are – probably translate into our having stronger moral reasons in general. The hope is that this would enable constructivism to yield a more impressive result along the vertical plane of extensional adequacy described above.

My second motive for wanting to explore a peak-experience-based idealisation is more theory-neutral in character. Even before we have considered constructivism or any particular theory of ethical truth, we might note that peak experiences have the phenomenology of *heightened insight*, of being morally and normatively insightful. (I will unpack this phenomenology in far more detail in §4.1 and §4.2; this is just a taster of what is to come.)

This fact – the fact that peak experiences seem, from the first-person point of view of the agent undergoing them, to communicate profound moral and normative insights – provides us with an independent reason for including them in our theory of value, *whatever that theory turns out to be in the end* (it need not be constructivist in flavour). In other words, we have reason, independent of constructivism, to be interested in peak experiences. And, as I hope to have shown in the preceding chapters, we have reason, independent of peak experiences, to be interested in constructivism. Enlightened constructivism meets these two motivations in the middle.

The experimental view that I am gesturing at here does not merely involve using peak experiences as an *epistemic tool* for discovering what the entailments of the practical point of view are; I am *not* proposing to wed a recognitionalist moral epistemology to a constructivist moral metaphysics. Rather, I want to explore a more

---

[27] For what it is worth, I personally do not find this result to be at all unintuitive, but I can see how for those in the meta-ethical moral realist camp it might seem like an unflattering aspect of the theory.

fundamental, meta-ethical change: tweaking the (Humean) constructivist method for generating truths about one's moral reasons.

As discussed in §3.1, this 'method' consists of nothing more than logical and instrumental entailment from a given agent's evaluative standpoint (i.e. the full set of attitudes of valuing held by that agent – in either an occurrent or dispositional form – together with any non-normative facts that have a bearing on those attitudinal valuations). Not to belabour the point, but, once again, *it is the entailments from within the practical point of view that constitute ethical/normative truth under Humean constructivism, as opposed to emergence from a correct hypothetical procedure*. Emergence from a correct procedure is not the truth-maker in this particular variant of constructivism. Humean constructivism does not make reference to any kind of procedure – even just as an epistemic tool or heuristic device. Rather, the central notion is of logical and instrumental entailment from the practical point of view (of a particular agent). I have been using the term 'method' as shorthand for this notion of entailment.

Enlightened constructivism involves tweaking the constructivist method. I will explain in §4.3 ('Idealising the Constructivist Method') precisely what sort of tweaks I have in mind here. Let me be clear from the outset, however, that I am *not* committed to the view that enlightened constructivism really is a plausible position – or even an improvement over unmodified, non-idealising (Humean) constructivism. My stance is exploratory rather than evangelising: I am not trying to convert people to the cause of enlightened constructivism; I am merely trying to explore and evaluate the view in the hope of rendering the overall package of meta-ethical constructivism more attractive and interesting.

Enlightened constructivism could be a promising theory, but if it turns out *not* to be – e.g. if the theory faces significant unanswered challenges, as I will argue §4.4 and §4.9 that it does – then we can retreat to the unmodified, non-idealising version of constructivism, which, as I hope to have shown in Chapters 2 and 3, still has much going for it. In short, if a peak-experience-based idealisation turns out to be unworkable and riddled with problems, then this does not undermine the broader constructivist approach, which remains promising.

With the preceding remarks in mind, the first three sections of this chapter, §4.1, §4.2, and §4.3, will focus on explaining how enlightened constructivism would work and on discussing the benefits that such a view would yield. In §4.4 I will introduce and evaluate some key objections to the position. My discussion of these objections will segue into a discussion of the first-person, experiential, phenomenological side of meditation practice (§4.5 onwards). I will explore whether certain forms of meditation, such as mindfulness meditation and loving-kindness meditation, can be thought of as reliable methods for eliciting ethically salient peak experiences. In §4.8 I will propose modifying the enlightened constructivist method such that instead of drawing on conventional peak experiences to construct truths about one's moral reasons, we draw on peak experiences that are elicited through the methodology of loving-kindness meditation (i.e. I will propose a modification *of* the modification). Ultimately, I will argue although enlightened constructivism represents an interesting position in logical space – and is an earnest attempt to incorporate the axiologically relevant phenomenon of peak experiences into a constructivist theory of value – the theory faces major unanswered challenges.

Before I proceed to the substance of this chapter, let me briefly reiterate a point made in the introduction to this thesis. The structure of this thesis more closely resembles an anthology of essays than an overarching narrative oriented towards the defence of a core set of claims. That is to say, the four chapters of this thesis are, to a large degree, free-standing and self-contained. This is especially true of this chapter, Chapter 4. The analogy I would draw concerning this chapter and its relation to the rest of the thesis is of a mitochondrion in a cell. Mitochondria are tiny organelles existing within human cells that produce the energy used by the cell (Harari 2011: 61). Mitochondria are originally descended from parasitic bacteria that were distinct organisms entirely, but at some distant point in our evolutionary past they merged in a synergistic way with the primordial precursors of human cells (Dawkins 1996: 132). In short, mitochondria are miniature, self-contained cells existing inside human cells. For instance, they 'have their own set of genes, which is completely separate from the DNA in the cell's nucleus' (Harari 2011: 61).

This chapter is like a mitochondrion that has merged with the cell that is this thesis. It started off as a distinct, parasitic bacterium, but now it is a synergistic organelle existing within the larger whole. I just mention this in case the contents of

this chapter seem somewhat digressive and only tenuously linked to the subject matter of the three chapters that have preceded it. This chapter can be thought of as a miniature thesis nested within the thesis: distinct and self-contained but also synergistic.

4.1 DEFINING PEAK EXPERIENCES

Now that I have set out my reasons for introducing the topic of peak experiences, let me turn to the all-important task of defining peak experiences. I think the best approach here is oblique rather than direct. Rather than baldly stating what peak experiences are, I will proceed by ostension. I will offer three examples of archetypal peak experiences taken from classic literature, and will then build up from there. The first example is from *War and Peace,* by Tolstoy. In this example, the Russian nobleman Andrei Bolkonsky is wounded while fighting the French, and this acts as a trigger for a peak experience:

'Hurrah!' shouted Prince Andrei, and scarcely able to hold up the heavy standard, he ran forward with full confidence that the whole battalion would follow him ... One soldier moved and then another and soon the whole battalion ran forward shouting 'hurrah!' and overtook him ... He heard the whistle of bullets above him unceasingly and to the right and left of him soldiers continually groaned and dropped. But he did not look at them: he looked only at what was going on in front of him – at the battery. He now saw clearly the figure of a red-haired gunner with his shako knocked awry, pulling one end of a mop while a French soldier tugged at the other. He could distinctly see the distraught yet angry expression on the faces of these two men, who evidently did not realise what they were doing. 'What are they doing?' thought Prince Andrei as he gazed at them. 'Why doesn't the red-haired gunner run away since he is unarmed? Why doesn't the Frenchman stab him? He won't get away before the Frenchman remembers his bayonet and stabs him ...' Prince Andrei did not see how it ended. It seemed to him as though one of the soldiers near him hit him on the head with a full swing of a bludgeon. It hurt a little, but the worst of it was that the pain distracted him and prevented his seeing what he had been looking at. 'What's this? Am I falling?

My legs are giving way', thought he, and fell on his back. He opened his eyes, hoping to see how the struggle of the Frenchmen with the gunners ended, whether the red-haired gunner had been killed or not, and whether the cannon had been captured or saved. But he saw nothing. Above him there was now nothing but the sky – the lofty sky, not clear yet still immeasurably lofty, with grey clouds gliding slowly across it ... 'How was it I did not see that lofty sky before? And how happy I am to have found it at last! Yes! All is vanity, all falsehood, except that infinite sky. There is nothing, nothing, but that ... Yes, I did not know anything, anything at all till now' (2010: 299–309).

I will hold off analysing this example until I have finished quoting the other two. The next example is taken from *Anna Karenina*, also by Tolstoy:

The servant on duty, who was not asleep, lit candles for him and was about to leave, but Levin stopped him. This servant, Yegor, whom Levin had not noticed before, turned out to be a very intelligent, nice, and above all kind man.

'So, do you find it hard staying awake, Yegor?'

'Can't be helped! It's my job. It's easier in a gentleman's house; but the pay is better here.'

It turned out that Yegor had a family – three boys and a seamstress daughter, whom he wanted to marry off to a clerk in a saddler's shop. This prompted Levin to share with Yegor his idea that the main thing in marriage was love, and that with love one would always be happy, because happiness only resides within oneself. Yegor listened attentively, and obviously fully grasped Levin's idea, but confirmed it in a way Levin was not expecting, by remarking that when he had been in service with good folk, he had always been happy with his masters, and was thoroughly happy with his master now, although he was a Frenchman.

'What a wonderfully kind man', Levin thought. 'And you, Yegor, when you married, did you love your wife?'

'Of course I did', replied Yegor.

And Levin saw that Yegor was also in a euphoric state, and on the point of divulging all his most intimate thoughts.

'My life has also been surprising. Ever since I was a child ...' he began with shining eyes, clearly infected by Levin's euphoria in the same way that people are infected by yawning (2014: 404).

The final example is from Dostoyevsky's dark masterpiece, *The Brothers Karamazov*:

His soul, brimming with ecstasy, was yearning for freedom, for wide-open spaces. Overhead, stretching into infinity, was the heavenly dome, full of silent, shimmering stars. From the zenith to the horizon stretched the forked outlines of the faintly visible Milky Way. A cool, silent, motionless night had enveloped the earth … Alyosha stood gazing; suddenly he fell to the ground, as though stunned.

He did not know why he was embracing the earth, he could not explain to himself why it was that he wanted to kiss it with such abandon, to kiss the whole of it, and yet he kept kissing it as he wept and sobbed, drenching it with his tears, and passionately swearing to love it, to love it for ever and ever … It was as though threads from all of God's countless words had converged in his soul … He wished to forgive everyone for everything and to ask forgiveness … with each passing moment he became distinctly, almost palpably aware that something as firm and immutable as the vault of heaven was entering his soul. An idea seemed to be taking possession of his mind … He fell to the ground a weak adolescent, but when he rose to his feet he was a hardened warrior for life, and he felt and recognised this in a flash of ecstasy (1994: 456–457).

Now that we have three examples of peak experiences on the table, we can begin by identifying some recurring motifs. The first is *intensity* (Privette 1983: 1361). Peak experiences are intense experiences in that they take a positive emotion, typically joy, and amplify it to an extraordinary degree (ibid.: 1363). The joy one experiences

is 'that of a thousand years crushed into a moment' (Schmitt 1994: 170). When joy reaches this fever pitch of intensity, it is probably more accurately described as 'euphoria' or 'ecstasy'. Indeed, the term 'euphoria' is used in the above extract from *Anna Karenina*; the term ecstasy is used in that from *The Brothers Karamazov*. The second recurring motif relates to duration. Peak experiences do not last very long. Of the three examples of peak experiences quoted above, none appear to last more than a few minutes. It is hard to imagine a peak experience lasting for twenty hours, say (unless one has some pharmacological assistance).

The third motif is a sense of heightened clarity and insight. When one is in the midst of a peak experience, one feels as if the scales have fallen from one's eyes. One is able to see reality in a whole new light. It is as though important truths that had hitherto been obscured are suddenly revealed, as when the clouds part to reveal a clear sky overhead. In the case of Prince Andrei Bolkonsky, in the first example I quoted, this is almost literally true. As Bolkonsky lies there on the battlefield, bleeding out from his wounds, with the sounds of war raging all about him, he gazes up at the 'immeasurably lofty' sky, and everything else seems to pale into insignificance (2010: 299–309). Many of the things that he previously thought were important – honour, military glory, defeating Napoleon, etc. – now seem small, mean, and trivial by comparison. Bolkonsky thinks to himself, 'All is vanity, all falsehood, except that infinite sky. There is nothing, nothing, but that ... Yes, I did not know anything, anything at all till now' (ibid.). During such experiences, it is as though one's mind has been infused with lucidity; one can see further, more clearly, and can make connections and grasp truths that one is generally not able to during episodes of mundane waking consciousness. As the Irish poet W. B. Yeats put it, 'our minds expand convulsively or spread out slowly like some moon-brightened image-crowded sea' (1968: 245).

The final recurring motif is that peak experiences usually have an ethical/ normative dimension in that they precipitate a significant change – sometimes a seismic upheaval – in one's moral and normative judgments (i.e. one's evaluations). We can think of a peak experience as being a bit like the eruption of a volcano. The tremors in the earth caused by the eruption are often strong enough to shake the very foundations of one's moral worldview. As Aaltola (2015: 289) puts it, peak experiences often have the power to 'spark foundational change in one's moral

outlook'. Just such a foundational change seems to have taken place in Alyosha, the character in the extract from *The Brothers Karamazov* quoted above. Alyosha is overcome by a feeling of rapturous euphoria. He falls to the ground 'as though stunned' and starts embracing the earth and sobbing convulsively (Dostoyevsky 1994: 456–457). He swears to love the whole earth forever. His emotion then takes on an other-regarding dimension: he wishes to 'forgive everyone for everything and to ask forgiveness' (ibid.).

It is at this point that Alyosha's peak experience seems to crystallise into a normative and ethical transformation. Alyosha becomes 'distinctly, almost palpably aware that something as firm and immutable as the vault of heaven [is] entering his soul'; an idea is 'taking possession of his mind' (Dostoyevsky 1994: 456–457). When he rises to his feet again, it is as a transformed being. He is no longer a 'weak adolescent' but a 'hardened warrior'. Lest the term 'warrior' convey the wrong impression here, it should be noted that Dostoyevsky meant a *warrior in the service of Christ*, a Christian warrior, an almost angelic being (Britannica Academic 2023: 'Fyodor Dostoyevsky'). Indeed, following this peak experience, Alyosha commits himself to acting in a highly other-regarding, self-denying, altruistic, and compassionate way.

One interesting facet of the ethical dimension of peak experiences is that the changes that occur in one's moral judgments during a peak experience are often impelled by a feeling of intense love. Research into real-life cases of peak experiences has shown that it is very common for people who have undergone such experiences to report having felt 'a generalized love toward everything and everybody and, notably, toward what is otherwise repulsive to them' (Schmitt 1994: 174).[28] I have already spoken about how peak experiences are primarily characterised by the emotion of very intense joy (I am taking the terms 'euphoria' and 'ecstasy' to be synonyms for this emotion). But it is worth noting that peak experiences do also very often involve a feeling of overflowing, boundless love, what we might call 'universal love'. Generally speaking, the feeling of intense joy comes first, and it then gradually takes on an other-regarding dimension and morphs into a

---

[28] The terms 'everything' and 'everybody' are somewhat nebulous here; I take the quote to mean that one feels a generalised love towards *all living things*. This could include non-human animals, such as any wild animals that one happens to observe while undergoing a peak experience.

feeling of universal love (which usually prompts changes in one's moral outlook). This is not to say, however, that these two emotions are mutually exclusive: during the segue from euphoria to universal love there may be a liminal period where the two feelings co-exist in a kind of synaesthetic harmony.

We can see an instance of this sort of overflowing, boundless love in the extract from *Anna Karenina* quoted above. When Levin the nobleman returns to his hotel late one night, he is brimming with ecstasy, and this feeling carries over very naturally into one of universal love. His feelings are too intense to keep bottled up inside him, so he lets them flow forth towards the nearest fit object in sight: the servant Yegor, who is working the night shift at the hotel. Levin engages Yegor in conversation. All class boundaries seem to melt away; the two men speak to each other from the heart. Interestingly, Levin's ecstatic mood seems almost contagious in that it spreads to Yegor, who begins recounting his life story with 'shining eyes' – until he is rudely interrupted by the sound of a bell summoning him elsewhere (Tolstoy 2014: 404).

Before I present a proper definition of peak experiences, let me add a quick word regarding the contexts in which such experiences most frequently occur. Research has shown that particular settings and activities act as powerful triggers for peak experiences: 'solitude, prayer, meditation, deep relaxation, physical accomplishment, and being in nature (particularly water, wild animals, sunsets, and mountains)' (McDonald 2009: 371). The most common trigger seems to be psychological isolation in relatively untouched and remote natural surroundings, such as 'the open sea, the wide shore, a bare windy mountain or hilltop, or a place where there are ruins—desolate places' (Schmitt 1994: 173). There is something about being in nature by oneself – like William Wordsworth wandering o'er hill and dale – that is conducive to the having of peak experiences (Bethelmy and Corraliza 2019; Aaltola 2015). One is far more likely to have a peak experience while walking alone through the woods at dusk than while sitting in a boring meeting at work surrounded by colleagues who are droning on about something or other. However, research in this area is yet to determine what precisely it is about being alone in nature that tends to induce such experiences in us (McDonald 2009: 371).

With the preceding remarks in mind, let me offer the following definition of peak experiences. Peak experiences are experiences of very intense joy – often combined with very intense love – that are short in duration and which involve a feeling of heightened clarity, insight, and higher-order stability on the part of the individual experiencing them; they generally have an ethical/normative dimension to them in that they allow one to see the possibility of a more coherent version of one's set of values and/or cause transformative changes in one's values; they are perspective-altering in that they grant one a seemingly more profound perspective on life, one which is rich with meaning and significance; they are sufficiently rare and extraordinary that they stand out in stark contrast to the experiences that surround them in time and space.[29] In characterising peak experiences primarily in terms of the emotion of joy, I am not suggesting that there is no cognitive component to them. One might, for instance, form new conceptual connections during a peak experience. One might even internally vocalise certain thoughts to oneself, as Prince Andrei Bolkonsky does while he is lying there on the battlefield, looking up at the sky: 'All is vanity, all falsehood, except that infinite sky. There is nothing, nothing, but that ... Yes, I did not know anything, anything at all till now' (Tolstoy 2010: 299–309).

Peak experiences are a focal-case concept (Owen 1960). There are clear and central cases of peak experiences that one can point to in an ostensional manner, as with the three examples from classic literature that I quoted earlier. But there are also less clear and less central cases – when exactly does an ascension become a full-fledged peak? Not much hinges on where *exactly* we draw the boundaries of the proper use of the term 'peak experience'. The central territory of the concept is not threatened by minor boundary disputes along its borders. No doubt, there are borderline cases of peak experiences, such as when one is flooded by a feeling of joy that is not quite intense enough to start shading into full-on euphoria/ecstasy. But the existence of borderline cases presupposes the existence of cases that are not borderline (Gaskin 2018: 2). And it is these non-borderline cases that are of interest here. There are no 'non-stipulative necessary and sufficient conditions for something's being [a peak experience]', and peak experiences fade out, around their edges, into 'relatively unexciting or small-scale phenomena like the merely striking or

---

[29] I will clarify certain aspects of this definition, such as the notion of 'higher-order stability', in the next section.

surprising moment' (Chappell 2022: 9). But this does not pose a problem for the peak-experience concept, just as it does not pose a problem for other focal-case concepts. For instance, there are no 'non-stipulative necessary and sufficient conditions for something's being a mountain, either, and the category of the mountainous typically fades out around its edges into literally small-scale phenomena', but that does not stop the 'geologist from studying mountains, nor the alpinist from climbing them' (ibid.).

The definition of peak experiences that I present above is stipulative; it is adequate for the purposes of this chapter, but I have also tried to make it representative of the existing literature on the topic. I would say that it is mostly extensionally adequate to the antecedent uses of the term 'peak experience', but I make no claim that it is the 'best' or authoritative way of defining peak experiences. With this caveat in mind, we can stipulate that if an experience possesses all of the key features that I mention in my definition, then that is sufficient for it to qualify as a peak experience. I am taking the conditions stated in the definition to be jointly sufficient. None of the conditions, however, are individually necessary. The reason for this is that I do not want the definition to become excessively rigid and inflexible. The subject matter we are dealing with – i.e. peak experiences – is inherently nebulous and ineffable, and I want to allow some flexibility in my definition so that an experience which diverged only very slightly from the stated conditions might still qualify as a peak experience (I will pick up on this point in §4.7). I am not planning on making any elaborate deductions based on this definition; the purpose of the definition is merely to clarify and elucidate the notion of peak experiences.

4.2 In What Way Are Peak Experiences Morally and Normatively Insightful?
Now that we have a definition of peak experiences on the table, let us drill down on the ethical, normative, and axiological import of such experiences (in the next section, §4.3, I will look at how we can draw on peak experiences to idealise the constructivist method for generating truths about moral reasons). During a peak experience one temporarily occupies a privileged epistemic position. What I mean by this is that one occupies a superior vantage point from which to judge whether certain things/options in the world are good or bad, right or wrong, better or worse, required or optional, worthy or worthless, valuable or disvaluable, etc. Peak

experiences involve a form of epistemic privilege when it comes to discerning what *matters*, what is of *value* in the world.

I should be clear that in making this claim, I am not tacitly appealing to a recognitionalist account of value. I am not suggesting that peak experiences provide a window unto a stance-independent moral reality: a moral reality that is 'not dependent on or grounded in the attitudes, either actual or hypothetical, of any agent or possible agent' (Behrends 2013: 486). As I discussed in Chapter 1, the constructivist view of things is quite the opposite. The constructivist maintains that reason-giving status, value-status, etc., are 'conferred upon things by us'; normativity is conferred on the world by the evaluations of conscious creatures, as opposed to being recognised in the world by them (Street 2010: 370–371). So, when I talk, for instance, about peak experiences illuminating aspects of moral reality that had hitherto lain obscured, this should be read in a metaphorical and figurative sense, not a literal one.

Under a constructivist view of things, then, what could a privileged epistemic position during a peak experience consist in, given that it could *not* involve seeing an attitude-independent realm of value more clearly? What it consists in is seeing the possibility of a more coherent version of one's set of values. As discussed in Chapter 1, there are many ways in which agents can suffer failures of coherence with respect to their own values. For example, agents can fail to notice how certain values are in tension with one another. They can fail to grasp the full implications of certain values. They can fail to keep in view all of the values that they do, in fact, hold. They can fail to scrutinise, re-evaluate, and update their values, such as values that were drummed into them on mother's knee and which they have always taken for granted (e.g. we can imagine a child in North Korea being taught from a young age the value of jingoistic patriotism and unquestioning devotion to the Dear Leader, and then going through much of their adult life without ever really questioning these values). During a peak experience, one is better able to recognise at least some of these failures of coherence and to remedy them.

Consider the following example. Imagine that Celine, an overworked, overpaid senior executive at a major corporation, decides to take a weekend break in the countryside. During this weekend getaway, she goes walking solo along some forest trails with the intention of clearing her head and getting some relief from the

stressful, fast-paced office environment that has become her stomping ground. Let us say that during one of these forest excursions, Celine undergoes a peak experience. To add a bit of colour to this example, consider the following remarks by Henry David Thoreau on the kaleidoscopic beauty of nature:

> Sometimes I rambled to pine groves, standing like temples, or like fleets at sea, full-rigged, with wavy boughs, and rippling with light, so soft and green and shady that the Druids would have forsaken their oaks to worship in them; or to the cedar wood beyond Flint's pond, where the trees, covered with hoary blue berries, spiring higher and higher, are fit to stand before Valhalla, and the creeping juniper covers the ground with wreaths full of fruit; or to swamps where the usnea lichen hangs in festoons from the white spruce trees, and toadstools, round tables of the swamp gods, cover the ground, and more beautiful fungi adorn the stumps, like butterflies or shells, vegetable winkles … Once it chanced that I stood in the very abutment of a rainbow's arch, which filled the lower stratum of the atmosphere, tinging the grass and leaves around, and dazzling me as if I looked through a coloured crystal. It was a lake of rainbow light, in which, for a short while, I lived like a dolphin (Thoreau 2018: 188–189).

So, we can imagine Celine standing there, dazzled by the light of a rainbow, and undergoing a powerful peak experience. During this experience, Celine might see clearly for the first time how certain of her values are in tension with one another. She deeply values having a brilliant, successful career and climbing to the summit of the corporate world, say, but she also deeply values having a healthy mind and body and a well-tended spiritual life. She now realises that working 70 hours a week in an office in London is not a means to the end of deep fulfilment and contentment in life but actually an impediment to that end. She realises that if one does not have the leisure time to actually enjoy the fruits of one's labour, then there is not much point in earning a stratospheric salary; one is merely sacrificing one's current happiness at the altar of some imagined future happiness, like a squirrel that hoards nuts in the bole of a tree with the intention of enjoying them at some unspecified future date, a date that seems to continually recede into the distance like a mirage. Celine realises

that being alone and deeply immersed in nature has done more to promote her well-being and flourishing than any success in her professional life ever has.

As noted in §4.1, one of the key aspects of the phenomenology of peak experiences is that when one is in the midst of them, the feeling is not that one's ethical intuitions and values have been deranged; rather, the feeling is one of *heightened* clarity and moral insight. This is evident in the way that participants in studies on peak experiences describe the peak experiences that they have undergone. For example, here are some quotes from participants in a study conducted by McDonald et al.: 'It was a deep sense of connectedness, of heightened awareness, of real relaxation, of a slowed passage of time'; 'I could truly *see* for the first time'; 'I had the sensation of sticking my head above the water I'd been living in entirely up to that point, and seeing a new world' (2009: 378–379). A recurring theme in self-reports of peak experiences is a feeling of 'transcendental clarity (a sense of gaining an epistemological meta-perspective)' (Aaltola 2015: 286). The most common metaphor that is used is of *seeing* more clearly, *seeing* things that one is usually not able to see during periods of mundane waking consciousness.

One might ask, what exactly *is it* that one is seeing more clearly? I have used the word 'insight' five times in this chapter so far in characterising the phenomenology of peak experiences; one might ask, insight into *what*? What does this phenomenological aspect of peak experiences actually correspond to? Different theories of value will answer this question in different ways. I will offer the constructivist answer to this question. To be clear, I am not claiming that the constructivist interpretation of the phenomenology of peak experiences is the *best possible* interpretation; I am merely claiming that it is a plausible interpretation.

The constructivist interpretation is that during a peak experience, one sees the possibility of a more coherent version of one's set of values; one's mind is opened to new possibilities in the (metaphorical) realm of value that one had hitherto been cognitively closed to. The insight one gains during a peak experience involves, for example, seeing for the first time how certain values are in tension with one another, grasping the full implications of certain values, becoming suddenly cognisant of values that had hitherto been implicit or buried in dark recesses of the mind, etc. To return to the example of Celine, as she stands there, surrounded by the kaleidoscopic beauty of the forest, she realises for the first time how she has been

unwittingly acting in such a way as to frustrate her own deepest values. Her experience has allowed her to gain insight into what she *really* wants out of life, what is most important to her, what is required of her if she is to live up to her own deepest values, what the trade-offs are between acting in accordance with certain values, etc.

In truth, there is a bit more to the constructivist interpretation of the phenomenology of peak experiences than this. Peak experiences sometimes cause substantive changes in one's values that go beyond just remedying failures of coherence in one's set of values. Indeed, research in this area has shown that following a peak experience, many express a 'willingness to alter their personal ethics even on a foundational level' (Aaltola 2015: 286). So, we can think of there beings two kinds of change that can occur in one's subjective motivational set during a peak experience: change in the direction of greater coherence and transformative change. The former involves seeing the possibility of a more coherent version of one's set of values and then actualising that possibility by altering one's values accordingly. The latter involves any substantive change in one's values that goes above and beyond this (e.g. the addition of a new value to an agent's motivational set). To be clear, I am not suggesting that these two kinds of change are mutually exclusive; peak experiences might involve both kinds. But generally speaking, I think that peak experiences mostly just involve change in the direction of greater coherence, with transformative change being somewhat rarer.

To make this notion of transformative change a bit more concrete, let us return to the example of Celine. Imagine that as a result of her peak experience in the forest, Celine becomes a fervent convert to the environmental movement and forms new values relating to protecting the environment and conserving areas of great natural beauty. These changes to her motivational set go beyond just increasing the coherence of her states of valuing; they represent a seismic upheaval in her moral outlook. And an important point here is that upheavals of this sort are higher-order stable: it is not just that they take place but that we *approve* of their taking place – or, at least, we approve of their *having* taken place once they have, in fact, taken place. This approval is, I believe, implicit in the feeling of 'transcendental clarity' and heightened moral and normative insight that I spoke about earlier (Aaltola 2015: 286). If one feels that the perspective one occupies during a peak experience is

transcendentally clear, insightful, etc., relative to the perspective that one occupies during periods of mundane waking consciousness, then this feeling is going to confer implicit self-endorsement on any substantive changes in one's values that occur as a result of that peak experience.

At this point, one might object that just because we subjectively *feel* that we are seeing whether certain things/options in the world are good or bad, right or wrong, better or worse, required or optional, worthy or worthless, etc., more clearly than ever before, does not mean that we actually *are* doing so. We are not subjectively incorrigible; the feeling of heightened clarity and insight that one has during a peak experience might just be pure delusion – it might be analogous to the feeling of heightened clarity, or pseudo-clarity, that a deranged conspiracy theorist feels when he thinks that he has finally glimpsed the truth behind the façade (e.g. Hillary Clinton is actually dead, and it was a body double that was campaigning in her place during the 2016 election).

We cannot conclusively rule out the possibility that the feeling of heightened clarity and insight that comes with peak experiences is delusory. However, I think the principle of Inference to the Best Explanation is appropriate here. We have data that needs explaining, which is that during peak experiences, people consistently feel a sense of heightened clarity and insight. One explanation of this data is that this feeling is delusory and misleading. In other words, an agent's *feeling* that he is occupying a position of heightened clarity and insight (when it comes to matters of value) is not a reliable indicator of whether he *is*, in fact, occupying a position of heightened clarity and insight. The problem with this explanation is that it seems to generalise beyond peak experiences. Is a feeling of heightened clarity *ever* a reliable guide to whether one actually is seeing things more clearly? For example, most people probably feel a sense of heightened clarity after having their morning coffee/ tea than before having it. But who is to say that this feeling is not also delusory? Indeed, how do we know that when one is feeling *least* clear and insightful that one is not actually at one's *most* clear and insightful? Perhaps one is actually at one's clearest and most insightful in the first 30 seconds after one wakes up in the morning, and any feelings of grogginess and mental stupor are just delusory and misleading.

This explanation is intuitively unsatisfying and leaves us with quite a strange picture of the human mind, one in which certain core feelings, no matter how strong, do not reliably track reality. I think a more probable and parsimonious explanation of the data is that people consistently feel a sense of heightened clarity and insight during peak experiences because they *actually do* occupy a position of heightened clarity and insight during peak experiences. Of course, this inference is not certain; inference to the best explanation 'cannot ever show that an explanation *has* to be true, but (at best) that it is most likely to be true' (Wilkenfeld and Lombrozo 2015: 1062). The inference could be undermined by the introduction of a 'sufficiently good alternative explanation' (ibid.: 1060). However, in the absence of a credible alternative explanation, I think it is reasonable to operate on the assumption that one's feeling of heightened clarity, etc., during peak experiences is indicative of one's occupying a privileged epistemic position in the way that I have described.

I think that a realist about value, such as a meta-ethical moral realist, would probably have some misgivings about the above line of reasoning. The realist might say something like: 'Well, look, you've explained how under constructivism, the position of heightened clarity and insight that one occupies during a peak experience corresponds to seeing the possibility of a more coherent version of one's set of values, seeing what a more coherent version of one's subjective motivational set would look like. That makes sense, because coherence is the sort of thing that gauged in a *stance-independent* way. For example, if, as a result of having a peak experience, an agent manages to iron out some of the logical inconsistencies that had lain concealed in her motivational set up till that point, then it really is true to say that, other things being equal, her motivational set has become *more coherent*. But that only covers one kind of change that can occur during a peak experience: change in the direction of greater coherence. What about the other kind of change, transformative change? If an agent undergoes a seismic upheaval in her moral outlook during a peak experience, one which feels from her own, first-person point of view like a huge improvement, how do we know that it *really is* an improvement? Under a constructivist view of things, what *actual* improvement could this felt sense of improvement correspond to?'

All that the constructivist can really say on this point is that the felt sense of improvement during peak experiences – the ones which involve *transformative*

change, I mean – corresponds to an actual and substantial change in one's existing ethical/normative judgments, which may involve the formation of entirely new ethical/ normative judgments. Recall that under constructivism, moral judgments are evaluations, which are conative states that provide their possessor with a goal and a motivation to pursue that goal (for the sake of simplicity, I will focus on the moral domain and put the normative to one side). Evaluations differ from ordinary beliefs in that they are intrinsically motivational. The emotion of intense joy – and often, of intense love – that is characteristic of peak experiences generally prompts a reconfiguration of one's evaluations. This could involve: (a) causing the formation of new evaluations; (b) increasing the strength of certain existing evaluations; (c) decreasing the strength of certain existing evaluations; (d) eliminating certain evaluations entirely; (e) some combination of (a) through (d).

Under constructivism, there is no *stance-independent* way of judging whether a transformative change in a given agent's moral outlook *really is* an improvement (i.e. whether a felt improvement corresponds to an actual improvement). Such judgments can only ever be rendered with tacit reference to a particular evaluative standpoint or combination of standpoints. One cannot step outside of one's evaluative standpoint – or of *any* evaluative standpoint – and occupy a view from nowhere from which to 'objectively' judge whether a given change in an agent's moral outlook is a change for the better or for the worse (Street 2008: 220; Street 2012: 49; Goldman 2006: 476). Such judgments implicitly rely on *other* evaluative judgments; it is other evaluative judgments which set the standard against which such judgments can be passed.

To return to the example of Celine, we can only determine in a stance-*dependent* way whether the change in Celine's moral outlook is a genuine improvement. So, for instance, if we were to ask King Charles III whether the change in Celine's moral outlook is a genuine improvement, he would, given that he has been a committed environmentalist his whole adult life, likely say, 'Yes, the new pro-environmental, pro-conservation orientation to Celine's moral outlook represents a significant and very real improvement to said outlook.' But if we were to ask a conspiratorial climate-change sceptic what he thinks of the change in Celine's moral outlook, he might give a different answer.

Despite this, I think that, on the assumption that the constructivist picture of ethical truth is correct, the higher-order stability involved in peak experiences gives us *prima facie* reason to privilege those values that one holds, in either an occurrent or dispositional form, during peak experiences in order to construct truths about one's moral reasons. Before I get into the details of how exactly this 'privileging' would work, it is worth acknowledging that many realists about value would object that the constructivist's notion of what constitutes an *improvement* in one's moral outlook during a transformative peak experience is far too relativistic and revisionist. It seems to open the door to the possibility of *perverse* peak experiences: peak experiences that change an agent's moral outlook in a way that seems, from his own, first-person point of view, like an improvement but which, from the point of view of other agents, seems like the exact opposite. This is precisely the objection to enlightened constructivism that I will consider in §4.4, so I won't say any more about it for now. Suffice it to say that I am cognisant of the objections that a realist about value would raise to the argumentative manoeuvres that have been made thus far.

4.3 IDEALISING THE CONSTRUCTIVIST METHOD

So, how might we modify the constructivist method to incorporate an element of idealisation centred around peak experiences? Recall that under the standard, non-idealising version of constructivism (i.e. Humean constructivism), the inputs into the constructivist method consist of: (a) the *actual* evaluations of agents; (b) any relevant non-normative facts; (c) the bare-bones requirements that govern practical reasoning (i.e. the principles of logical entailment and means-end coherence). Under enlightened constructivism, we tweak one of the inputs into the method. Instead of saying that truths about an agent's moral reasons are determined by what follows, as a purely logical and instrumental matter, from the full set of that agent's *actual* evaluations in combination with any relevant non-normative facts, we say that truths about an agent's moral reasons are determined by what follows, as a purely logical and instrumental matter, from *the full set of actual evaluations that the agent in question holds during a peak experience*, plus any relevant non-normative facts.

Rather than constructing ethical truth using any old evaluations, we are using those evaluations that one holds – in either an occurrent or dispositional way – during a peak experience. It is as though we are taking a freeze-frame of one's

evaluations at a particular moment in time and marking that freeze-frame as the high-water mark of one's subjective motivational set. The attitudes that are captured in the freeze-frame then go towards fixing ethical truth (i.e. truths about one's moral reasons). This raises the question of *which* peak experiences we should be drawing on here; an individual might experience dozens or even hundreds of peak experiences throughout his life. It might be that the states of valuing that are revealed or elevated during one peak experience are at odds with those that are revealed or elevated during another. I think the way to get around this issue is to draw on whichever peak experience represents the highest peak in the mountain range that is one's life; we should go with the view from the highest summit, as it were. We can measure height using the feeling of transcendent clarity, insight, and higher-order stability that I spoke about in §4.2. One's 'highest' peak experience – at least for the purposes of this chapter – is that which, on balance, is the *most* joyous and which involves the *greatest* feeling of clarity, insight, and higher-order stability (higher-order stability with respect to any changes in one's evaluations that occur during a peak experience, whether these be changes in the direction of greater coherence or transformative changes – or both).

But *why* should we opt to restrict the constructivist method in this way? The reason for doing so is that the perspective we temporarily inhabit during a peak experience feels, from our own, first-person point of view, like a better, wiser, and deeper perspective than that which we occupy during periods of mundane waking consciousness, and this feeling carries with it an implicit self-endorsement of those states of valuing that are revealed, elevated, etc., during such an experience. Here is another way of looking at it. Let us assume for the sake of argument that (Humean) constructivism is true: truths about one's moral reasons – and about one's normative practical reasons more generally – are determined by what follows, as a purely logical and instrumental matter, from the full set of one's actual evaluations in combination with any relevant non-normative facts. Now imagine that one could step back and take a bird's-eye view of one's subjective motivational set over the course of one's life. Imagine that one could see which evaluations one held at any particular moment in time (up till the present). Let us say that one is offered a choice between two options. Option (a) is that one's moral reasons could be determined, at least in part, by the evaluations that one holds – in either an occurrent or dispositional form –

at the current moment in time. This is the case under the standard, non-idealising version of constructivism. Option (b), by contrast, is that one's moral reasons could be determined, at least in part, by those evaluations that one held during the highest peak experience of one's life (highest in the sense outlined in the preceding paragraph).

Phrasing it more directly, the enlightened constructivist might ask, 'Would you rather that your moral reasons were anchored to whatever evaluations you happen to hold right now, or that they were anchored to those evaluations that you held during the most intense peak experience you have ever undergone in your life? In other words, who do you want to determine your moral reasons: *you* now, or *you* when you were at the highest ebb of your experience, when your mind was as clear, expansive, loving, euphoric, etc., as it has ever been? Who is a better judge of what truly matters in life: *you* now or *you* when you were riding the crest of the highest wave of positive emotion that you have ever experienced?' I think that most people would agree that *on the assumption that constructivism is true*, it would be preferable to have one's moral reasons anchored to the subjective motivational set that one had during one's moment of greatest clarity and moral insight, as opposed to whatever subjective motivational set one happens to have at this current moment in time.

One problem with this view is that it would seem to render one's values static, frozen in time. One takes a freeze-frame of one's values at a particular moment in time – the moment of one's highest experiential peak – and then uses that freeze-frame to, in part, determine the nature of one's moral reasons. This does not seem to allow for a natural evolution in one's values depending on changing life circumstances. Presumably, we want to allow for *some* flexibility in our values across time. We do not want our values to be like slabs of granite, totally immovable. The way that the enlightened constructivist can get around this is by introducing a further layer of complexity into the idealisation. Instead of asking what one *in fact* valued during one's most intense peak experience, we can ask what one *would* value right now if one were, once again, in the state of mind that characterised one's most intense peak experience.

We can think of there being three relevant versions of oneself: (a) one's actual self at the current moment in time; (b) one's actual self at the moment of one's

highest peak experience; (c) a hypothetical version of oneself who has all of the same information, autobiographical memories, etc., as one's current, actual self but who inhabits the same extraordinary state of mind that one did during one's most intense peak experience. We can imagine this hypothetical version of oneself as *permanently* occupying the same lucid, expansive, euphoric, and loving perspective that one *temporarily* occupied during one's highest experiential peak. I will refer to this hypothetical version of oneself as one's 'peak self'. Under enlightened constructivism, then, one's moral reasons are determined, at least in part, by the hypothetical evaluations that one's peak self *would* hold if he or she were in one's present situation. One can outsource the determination of one's moral reasons to a hypothetical version of oneself who represents what one *would* be like if one were in a constant state of heightened clarity and moral insight – or, at least, what *seemed* very strongly like a state of heightened clarity and moral insight from one's own point of view when one was in it.

I think that this rather abstract notion of the 'peak self' is best illustrated by way of an example. The example I will use is taken from Book One of Homer's *Iliad*, wherein the great warrior Achilles gets into a heated quarrel with Agamemnon, who is the field marshal of the Greek forces. Before I quote the relevant extract, I should say that I am *not* proffering this as an example of a peak experience; I gave some examples of peak experiences taken from classic literature back in §4.1. Rather, the purpose of this example is to give a sense of why we, as actual agents, should care about the perspectives of our peak selves.

> But Achilles, the fast runner,
> glowered from beneath his brows and said: "Shameless fool!
> Greedy, how now can your speech gladly persuade any
> of the Achaeans either to go on an ambush or to fight
> in the hand-to-hand? *I* at least did not come here to war
> because of the Trojan spearmen. They have done nothing
> to me. They have not taken my cattle, nor horses.
> Not in Phthia with its very rich plowlands,
> the nurse of men, have they laid waste the harvest.
> For many are the shadowy mountains that lie between us,

and the echoing sea. No, we followed you—you dog
without shame!—that you might be happy, that you
might win honor from the Trojans for Menelaos—
and for yourself! Dog! You don't ever think about that,
do you? Are you troubled by that? You threaten yourself
to take *my* prize, for which I labored sorely?
The sons of the Achaeans gave her to me. I never
get a prize equal to yours, when the Achaeans take some
populous city of the Trojans, though I myself
bear the hard brunt of the fighting. When the spoils
are divided, your prize is always bigger. *I* get
some small, little darling thing and slouch off
to my ships, worn out with the war.
    "Okay, I'm off
to Phthia. I think it's better to head away in my beaked ships
than to hang around here and pile up endless wealth
for you, while I remain without honor!"

    The king over men,
Agamemnon, then said in reply: "Go then, if that's what you
want to do. Don't stay on my account. There are plenty who
will honor me, and Zeus above all, whose wisdom is great.
You are most hateful to me of all the god-reared chieftains!
You ever love contention and war and battle. If you are strong,
that is because some god gave you the gift. Go home with
your ships and your companions. Rule over the Myrmidons!
I don't like you, I don't care if you are angry.
    "But I tell you what
I'm going to do. Just as Phoibos Apollo takes away Chryseïs—
and I'll send her in a ship with an escort—even so
I will myself come and take the high-cheeked Briseïs,
your prize, that you might know how much stronger I am
than you, and so that any other may think twice

before saying that he is my equal and liken himself to me
to my face!"

So he spoke. A great hurt arose inside
the son of Peleus, and his heart within his shaggy
breast was divided in two ways, either to draw his sharp
sword from his thigh, break up the assembly, then kill
the son of Atreus, or whether he should stop his anger,
bridle his tumult. While he pondered in his heart and in his spirit,
he drew out the great sword from its scabbard, but Athena came
from heaven—the goddess Hera of the white arms sent her
because she loved and cared for both men equally.

Athena stood behind him and she seized the son of
Peleus by his light-colored hair. Only he could see her.
He was amazed. He turned around, and right away he recognized
Pallas Athena. Her eyes shone with a terrible light. He spoke
to her words that went like arrows: "Why have you come,
daughter of Zeus who carries the goatskin fetish?
To see the insolence of Agamemnon, son of Atreus?
But I will tell you this, and I think it shall come
to pass—through his insolence he will quickly lose his life!"

Flashing-eyed Athena answered him: "I have come
from heaven to stop your anger, if I can persuade you.
White-armed Hera sent me, because she loves
and cares for both of you. So come, let go
of this contention. Unhand your sword. Abuse him
with words. Tell him how things will be. For I promise you
that this will come to pass: You shall one day have
three times as many gifts because of this violence.
Hold back, trust us."

Achilles, the fast runner, answered:
"What can I do but obey the two of you, goddess,

even though I am seething? It is better that way. The gods

listen to him who obeys." And he stayed his heavy

hand on the silver sword-hilt, and back into the scabbard

he thrust the great sword, not disobeying the word of Athena.

She went off to Olympos to the house of Zeus who carries

the goatskin fetish, to be with the other gods (Powell 2013: 45–48).

So, Achilles faces an urgent moral dilemma: should he take immediate vengeance by slaying Agamemnon, or should he 'bridle his tumult' (ibid.: 46). We might say that the inner struggle that goes on in Achilles' mind is derivative of a tension in his subjective motivational set. This is to embellish the example, but let us say that Achilles' evaluations are such that he has some normative practical reasons to slay Agamemnon, but these reasons are in tension with his moral reasons to refrain from slaying Agamemnon. To be clear, I am *not* claiming that this is actually a plausible reading of Homer's text; I am not seriously attempting to derive any moral truths from the text of *The Iliad*. Rather, I am using the example of Achilles for purely illustrative purposes: to add a bit of colour to my black-and-white portrait of enlightened constructivism.

Recall that under constructivism, it is not *only* one's evaluations that go towards determining one's reasons; as discussed in §3.8, the relevant non-normative facts have an important role to play as well. But so as not to over-complicate this particular example, let us stipulate that the relevant non-normative facts are such that Achilles really does have normative reasons to slay Agamemnon and moral reasons to refrain from slaying him.

Just for some historical and mythological context, it is worth noting that Achilles was both a prince and a demigod (Britannica Academic 2023: 'Achilles'). He was the son of Peleus, who was the king of the Myrmidons. His mother was the divine sea-nymph Thetis. Achilles was famed as the greatest warrior of the Greek army. The reason I mention this is that to the modern reader, the above extract from *The Iliad* might seem almost parodic, a parody of macho posturing. It may be hard to imagine Achilles *actually* having a normative reason to slay Agamemnon. But in the brutal

world of Bronze Age warlords, one's reputation could be made or broken depending on how one handled a quarrel of this sort. In the heated quarrel with Agamemnon, who is the commander-in-chief of the Greek forces in the siege of Troy, it is not just Achilles' personal honour and reputation that is at stake. Achilles is the representative of his father's kingdom abroad, and he is the chieftain of the Myrmidon warriors, who form the fiercest contingent of the Greek army at Troy. So, the reputation of his kingdom and that of his warriors is also at stake in the quarrel with Agamemnon. Agamemnon treats Achilles in a highhanded and domineering way, as a king treats a subject. But Achilles is not a subject: he himself is heir to a kingdom. With this in mind, one can see how it is not such a simple matter for Achilles to 'turn the other cheek' or 'swallow his pride' in this dispute.

We can imagine Achilles holding the following evaluations. 'It is required of me that I defend the honour and reputation of my father's house and that of my warriors. It is better to die doing one's martial duty in the heat of battle than to live a long life as a coward. To be worthy of my reputation as the greatest living warrior among the Greeks, I must never be seen to ignore an insult or to hesitate out of fear. When I act, I act not just as one man but as the embodiment of my kingdom and with the eyes of all my men upon me.' And so on. Obviously, this is just a vignette; I am not going to exhaustively enumerate all of Achilles' evaluations. But hopefully this is enough to get a sense of what sort of motivational set Achilles might have had. We can see how, depending on what the relevant non-normative facts were, these evaluations might have generated strong normative practical reasons for Achilles to slay Agamemnon.

However, this is only one half of the conative story. We can imagine Achilles also having certain evaluations that are more pro-social, cooperative, kindly, and other-regarding in nature. *The Iliad* does not portray Achilles as some callous ruffian who is totally indifferent to the norms of morality. Rather, Achilles is the archetype of the valiant warrior. Although he slays men in the course of battle, he does not shed blood indiscriminately. He does not cruelly slaughter women or children during the sacking of a town, for instance. He follows a kind of Bronze Age warrior code and seems to place a lot of emphasis on observing certain norms and customs. For example, when Achilles' close friend, Patroklos, is killed while fighting the Trojans, Achilles stages

elaborate funeral games in his honour and, in so doing, shows a softer side to his character:

> So he spoke, and all together they raised their sorrowful wail,
> and Achilles was their leader. Three times they drove their horses
> with fine manes around the dead body, keening. Among them
> Thetis stirred the desire for weeping. The sands were moistened,
> the armor of the men was wet with their tears. They grieved
> for their mighty deviser of rout (Powell 2013: 515).

Let us say that Achilles' more other-regarding evaluations generate moral reasons for him to act in certain ways. To bring this back to the quarrel with Agamemnon, let us say that Achilles has moral reasons to refrain from drawing his sword and slaying Agamemnon right there in front of the chieftains of the Greek army; to do so would violate certain norms that Achilles himself values (e.g. showing the appropriate degree of respect, in a military context, to the field marshal of the army for which one has pledged to fight; avoiding the use of lethal force, outside of the battlefield, against those who are not themselves trying to do one any physical harm; not sowing discord among the Greeks in such a way as would benefit the enemy against which they are fighting, etc.). Again, I am embellishing the original example, but this is all with an eye towards showing how an idealising form of the constructivist method centred around peak experiences would work.

So, Achilles' reasons are in tension. He is vacillating between slaying Agamemnon or not slaying Agamemnon. And it is at this exact moment that bright-eyed Athena comes down from Mount Olympus to offer him wise counsel. She bids him to cool his fury, to refrain from acting rashly. She urges him to 'let go / of this contention. Unhand your sword. Abuse him / with words. Tell him how things will be … Hold back, trust us' (Powell 2013: 48). Achilles heeds her advice: he puts his sword back in its sheath. But *why* does Achilles heed her advice? To offer a constructivist interpretation, it is because he knows that Athena, being a goddess, occupies a wiser and more far-seeing ethical/normative perspective than he does. Athena sees the situation – his quarrel with Agamemnon and its potential ramifications – with a degree of clarity and insight that surpasses his own.

To tie this back to our earlier discussion of enlightened constructivism, we can view the goddess Athena as a metaphor for one's *peak self*. Recall that one's peak self is a hypothetical version of oneself who has all of the same information, autobiographical memories, etc., as one's current, actual self but who inhabits the same extraordinary state of mind that one did during one's most intense peak experience. One's peak self is not an *ideally* coherent, informed, or imaginative being; it is just a hypothetical version of oneself that *permanently* occupies the same lucid, expansive, euphoric, and loving perspective that one *temporarily* occupied during one's highest experiential peak. In this sense, enlightened constructivism differs significantly from other 'ideal observer' versions of ethical subjectivism, which usually stipulate that the hypothetical beings whose mental states go towards fixing ethical truth are in possession of 'full knowledge of the non-moral facts' and think 'impartially, dispassionately, and consistently' (Sinclair 2020: 43).

For instance, Firth (1952) suggests that an ideal observer must be 'omniscient with respect to non-ethical facts … no limits are put on the kinds or the quality of factual information which are available to influence his ethically-significant reactions' (334); he must also be 'omnipercipient' in that he must be able to simultaneously 'visualize all actual facts, and the consequences of all possible acts in any given situation, just as vividly as he would if he were actually perceiving them all' (335); and he must be dispassionate in that he is 'incapable of experiencing … such emotions as jealousy, self-love, personal hatred, and others which are directed towards particular individuals as such' (340).

Under enlightened constructivism, there is an experiential bridge connecting one's actual self to one's hypothetical, peak self: one has, for a short period of time, experienced the very same lucid, expansive, joyful, loving, etc., state of mind that we are stipulating that one's peak self perpetually enjoys. One was, for a short period, *identical* to one's peak self. There is no such connecting thread under other ideal observer theories, such as Firth's (1952). For instance, no *actual* agent has ever experienced complete knowledge of the non-moral facts or attained ideal coherence, imaginativeness, etc. – not even for a short while. Under such theories, there is an unbridgeable experiential gap between actual agents and the hypothetical beings whose mental states go towards fixing ethical truth; such beings are almost like alien life-forms, somewhat reminiscent of Spock from *Star Trek*.

So, what does enlightened constructivism ask of us? It asks the following. In the same way that Achilles defers to Athena because of the privileged epistemic perspective that she occupies relative to him, so, too, we should defer to our peak selves because of the privileged epistemic perspectives that they occupy relative to us. What I mean by 'defer' is that one should 'take on' or acquire those moral reasons that one's peak self would have if he or she were in one's current position. Let me formulate this claim more precisely. Under enlightened constructivism, *X* is a moral reason for *A* (e.g. Achilles) to *Y* iff *X* is logically and instrumentally entailed from within the evaluative standpoint of *A*'s peak self at a particular moment in time, the evaluative standpoint being composed of the *full set* of hypothetical attitudes of valuing held by *A*'s peak self – in either an occurrent or dispositional form – together with any non-normative facts that have a bearing on those attitudinal valuations.

But why should one concern oneself with the moral reasons that are entailed from within a *hypothetical* evaluative standpoint? Why not just focus on those moral reasons that are entailed from within one's *actual* evaluative standpoint? Speaking metaphorically, why should Achilles listen to Athena? The reason is that the state of mind that one's peak self perpetually enjoys is a state of mind that *seemed* – when one briefly occupied it during one's highest peak experience – like a state of extraordinary clarity and insight relative to one's everyday perspective. And if one *could* enjoy that state of extraordinary clarity and insight in perpetuity – i.e. if one could transform one's whole life into a never-ending peak experience – one would choose to do so. We should defer to the ethical perspective of our peak selves because such beings occupy a perspective that we know from direct, first-person experience to be a wiser and more far-seeing perspective, ethically speaking.

As discussed earlier, peak experiences generally alter our evaluations by strengthening those which are more pro-social, altruistic, and other-regarding in orientation. This is evident, for instance, in the extract from *The Brothers Karamazov* that I quoted in §4.1. In this extract, the character 'Alyosha' is struck as if by a lightning bolt of ecstatic love, and as he feels that emotional electricity coursing through his body, it jolts his subjective motivational set in a more ethical and other-regarding direction. Alyosha falls to the ground 'as though stunned' and starts embracing the earth and sobbing convulsively (Dostoyevsky 1994: 456–457). He

swears to love the whole earth forever. He wishes to 'forgive everyone for everything and to ask forgiveness' (ibid.). It seems probable that, depending on the relevant non-normative facts, such an experience as Alyosha undergoes would strengthen one's moral reasons. In the case of Alyosha specifically, he commits himself, following his peak experience, to behaving in a highly other-regarding, self-denying, altruistic, and compassionate way, such as by volunteering to teach the children in the local orphanage and by setting aside a meagre living allowance for himself and donating the rest of his money in alms to the poor.

Under enlightened constructivism, then, one's moral reasons are anchored to the hypothetical evaluations that one's peak self would have if he or she were in one's current position. One's peak self will, generally speaking, have a set of evaluations that is considerably more other-regarding in orientation. Our peak selves are like a holographic projection of us at our best; they represent who we could be if we always lived from that place of heightened clarity and insight that we temporarily inhabit during our most expansive moments. When we outsource the determination of our moral reasons to our peak selves, we come into possession of those moral reasons that we *would* have if we were always in the most lucid, joyful, and loving state of mind that we have ever experienced.

4.4 OBJECTIONS TO ENLIGHTENED CONSTRUCTIVISM

Now that I have sketched out how enlightened constructivism would work, let us turn to some objections to the view. The first objection is epistemic in nature: how can one know what evaluations one's peak self would hold? One's peak self is a hypothetical being with hypothetical evaluations. If one cannot know what these attitudes are, then it would seem that one cannot know what one's moral reasons are. If, as the enlightened constructivist claims, truths about one's moral reasons are anchored to the hypothetical evaluations of one's peak self, then this would seem to provide us with 'unattainable, unknowable standards' of correctness in the moral domain (Kawall 2006: 361).

The enlightened constructivist response to this is to reiterate a distinction I first broached in Chapter 1 when discussing the standard, non-idealising version of (Humean) constructivism. Just like the regular constructivist method, the enlightened constructivist method is a *hypothetical* method: it need not be something that actually

happens. Enlightened constructivism is not so much a method of answering moral questions as it is a claim about the status of ethical truth. The position is not a decision-procedure; it is a thesis about what makes statements about moral reasons true. As such, the view does not claim that we are always in an epistemic position to know what the outputs of the constructivist method will be. There may be practical difficulties involved in actualising the method. It is not as though one can simply sit down with a pencil and paper and, after twenty minutes of intense cogitation, know exactly what the truths about one's moral reasons are.

But we should not mistake *no answers in practice* for *no answers in principle*. Just because it could, in practice, be very difficult to work out what the truths about one's moral reasons are under enlightened constructivism does not mean that there are *no* such truths to be known – or that we cannot ever know them. By analogy, consider a question like 'How many people on planet Earth are playing golf at this very moment?' Obviously, there would be enormous practical difficulties involved in figuring out the answer to this question. But there is no doubt that the question *has* an answer. The answer is just an integer (e.g. exactly 52,512 people are playing golf at this very moment). As discussed in Chapter 2, it seems like a plausible intensional constraint on any moral theory that whatever moral truth is, it cannot be something that is, *in principle*, completely inaccessible to us. Enlightened constructivism does not claim or imply this. It is not as though the domain of moral truth is walled off from us like some hidden, enclosed garden that no passerby can ever enter. There is no law of the universe preventing us from working out what our moral reasons are. But such knowledge would likely be difficult to access, like a very tough nut that one has to crack open before one can access the sweet kernel inside.

The reason that enlightened constructivism is able to satisfy the above intensional constraint – i.e. that whatever moral truth is, it cannot be something that is, *in principle*, completely inaccessible to us – is that the hypothetical beings whose attitudes go towards fixing ethical truth are not that far removed from us. The degree of idealisation involved in enlightened constructivism is nowhere near as extreme as in other ideal observer views, such as Firth's (1952), which I discussed in §4.3. Recall that under enlightened constructivism, one's peak self is a hypothetical version of oneself who has all of the same information, autobiographical memories, etc., as one's current, actual self but who inhabits the same lucid, expansive, joyful,

loving, etc., state of mind that one did during one's most intense peak experience. What this means is that there is an experiential bridge connecting one's actual self to one's peak self. One has, for a short period of time, experienced the same extraordinary state of mind that we are stipulating that one's peak self enjoys in perpetuity. One was, for a short while, *identical* to one's peak self.

By contrast, no such experiential bridge is involved in Firth's (1952) ideal observer theory, for instance. As I touched on in §4.3, Firth claims that the ideal observer whose attitudes go towards fixing ethical truth must be both omniscient and omnipercipient, among other things (1952: 334ff.). One might reasonably worry that under such a view as Firth's, moral truth is not just *practically* inaccessible to us but inaccessible to us *in principle*. As Kawall (2006: 361) puts it, 'How are we, as mere humans, ever to know what is morally right on an ideal observer theory [like Firth's]? We are not omniscient; nor do we possess unlimited reasoning abilities. The perspective of an ideal observer seems distant and removed at best.' We cannot take up the position of an omniscient being or even begin to anticipate what such a being's reactions would be in a given situation (Sayre-McCord 1994: 218). We cannot even dimly imagine what it would be like to be omnipercipient in the sense that Firth describes (i.e. to be able to simultaneously 'visualize all actual facts, and the consequences of all possible acts in any given situation, just as vividly as he would if he were actually perceiving them all' (1952: 335)).

But these concerns do not carry over to enlightened constructivism. Our peak selves are not omniscient, omnipercipient, etc. They do not hold any superlative qualities of this kind. As such, there is no 'devastating epistemic gap' between the perspective of one's peak self and one's actual self (Kawall 2006: 364). Virtually everyone will have undergone a peak experience at some point in their lives – though the intervals between such experiences can be quite long (e.g. it may be that one has, on average, a peak experience every nine months, say). I do not think it is beyond the capacity of the human imagination to imagine a hypothetical version of oneself – almost like a holographic projection of oneself – who permanently occupies the same lucid, expansive, joyful, loving, etc., state of mind that one temporarily occupied during one's most intense peak experience. We can imagine adding something to one's morning tea or coffee that turns one's life into one long, never-ending peak experience, almost as though one were a character in a musical that

never ends, a character prone to break into spontaneous song and dance at any moment of the day.

As already mentioned, enlightened constructivism is not a decision-procedure. I am not trying to show that ethical truth would always be *practically* accessible to us in every situation. I am merely trying to show that the perspective of our peak selves is not so far removed from that of our actual selves as to render 'the realm of morality … forever closed to us' (Kawall 2006: 362).

Let us consider a second objection to enlightened constructivism. Before I get to the objection itself, let me, as an *aide-mémoire*, briefly restate the definition of peak experiences that I presented in §4.1. Peak experiences are experiences of very intense joy – often combined with very intense love – that are short in duration and which involve a feeling of heightened clarity, insight, and higher-order stability on the part of the individual experiencing them; they generally have an ethical/normative dimension to them in that they allow one to see the possibility of a more coherent version of one's set of values and/or cause transformative changes in one's values; they are perspective-altering in that they grant one a seemingly more profound perspective on life, one which is rich with meaning and significance; they are sufficiently rare and extraordinary that they stand out in stark contrast to the experiences that surround them in time and space. As mentioned in §4.1, this is a stipulative definition aimed at clarifying and elucidating the principal subject matter of this chapter: peak experiences. I am not planning on making any elaborate deductions based on this definition, so we need not haggle too much over the details.

I argued in §4.2 that it is the sense of heightened clarity, insight, and higher-order stability during a peak experience that confers a special status on those evaluations that one holds – in either an occurrent or dispositional form – during such experiences. What the phenomenological dimension of heightened clarity and insight corresponds to, under constructivism, is that during a peak experience one sees the possibility of a more coherent version of one's set of values; the 'insight' involved is insight into what a more coherent version of one's subjective motivational set would look like. I elaborated in §4.2 on what this coherence might consist of (e.g. seeing, for the first time, how certain values are in tension with one another; grasping the full

implications of certain values; becoming suddenly cognisant of values that had hitherto been implicit or buried in dark recesses of the mind, etc.). During a peak experience, one is better able to recognise at least some of these failures of coherence and to remedy them. This accounts for one kind of change in one's values that may occur during a peak experience: change in the direction of greater coherence.

There is another kind of change that sometimes – though not always – occurs during a peak experience. This is transformative change: a major, substantive change in one's values that goes above and beyond merely increasing the coherence of one's set of values. This kind of transformative change might involve, for instance, the addition of an entirely new value to one's motivational set. I argued in §4.2 that there is a higher-order stability involved in changes of this kind during a peak experience: it is not just that such changes take place, but that we *approve* of their taking place – or, at least, we approve of their *having* taken place once they have, in fact, taken place. On the assumption that the constructivist picture of ethical truth is correct, I think that this higher-order stability gives us *prima facie* reason to privilege those values that one holds, in either an occurrent or dispositional form, during peak experiences in order to construct truths about one's moral reasons. I explained at length in the preceding section how exactly this 'privileging' would work. To be clear, I am not suggesting that these two kinds of change – change in the direction of greater coherence and transformative change – are mutually exclusive. Peak experiences could involve changes of both kinds. But generally speaking, I think that they mostly just involve change in the direction of greater coherence, with transformative change being rarer.

The first kind of change mentioned above, change in the direction of greater coherence, is the sort of thing that can be measured – at least *in principle* – in an 'objective', stance-independent way. That is, we can measure it with reference to standards of coherence that are *independent* of the evaluations of any particular agent. For example, if two particular evaluative judgments are in logical or practical contradiction with one another, then this fact – the bare fact of their being in contradiction – does not *itself* obtain in virtue of the contingent attitudes of any particular agent. But the same cannot be said of the second kind of change: transformative change. There is no stance-independent way, under the constructivist

162

view of things, of judging whether or not a radical, substantive change in an agent's values during a peak experience constitutes a genuine improvement in that agent's motivational set.

This gives rise to the following problem. We can think of peak experiences that substantively change an agent's values in a way that seems, from his own, first-person point of view, to have significantly *improved* his values but which, from the point of view of other agents, seems to have achieved the exact opposite. That is, a peak experience could change an agent's moral outlook in a way that is higher-order stable (from that agent's perspective) but which seems utterly perverse and deranged (from the perspective of other agents). To illustrate this point, consider the following quote from the great Mongol Emperor Genghis Khan: 'The greatest joy a man can know is to conquer his enemies and drive them before him. To ride their horses and take away their possessions. To see the faces of those who were dear to them bedewed with tears, and to clasp their wives and daughters in his arms' (Gat 2006: 427).

Assuming that this is a sincere communication from Genghis Khan, it licenses us to make certain inferences about the kinds of experiences he probably had throughout his life. In the course of conquering vast swathes of the known world, he likely had many experiences that had all the features listed in my above definition of peak experiences. Which is to say, Genghis Khan probably had many peak experiences in a military context. We can imagine him, for instance, standing on a hilltop and gazing out over the smouldering remains of a city he had just sacked and being flooded by a feeling of preternatural joy and clarity. It is not hard to imagine such an experience seeming, *from his point of view*, to communicate profound moral and normative insights. As he becomes overwhelmed by joy, the incomparable joy of conquest, Genghis Khan might think to himself: *Yes, this is exactly what I was born to do. I am fulfilling my destiny. I am a father to the Mongol people, and I must forge a new realm for them. To do this, I need to topple the existing order, to raze cities to the ground and enslave their citizens, to bend proud nations to my will. There can be no room for pity or weakness. If I am to walk the path that fate has cut out for me, if I am to build a new world for my people, I must first destroy the existing one. There will be terrible suffering, but it will all be for the good in the end.*

Most people would be horrified by the normative and ethical implications that Genghis Khan draws from his peak experience in the above vignette. Intuitively, we feel that anyone who judges himself to have moral reasons to engage in ruthless, genocidal military conquest is simply mistaken, and this is true even if the person in question has arrived at these judgments via an experience of transcendent joy, clarity, insight, etc. In other words, we feel that Genghis Khan's set of values has *not* been improved by this peak experience, and this is true no matter how intense Genghis Khan's subjective feeling of heightened clarity, insight, and higher-order stability was while he underwent the experience.

Unfortunately, the enlightened constructivist has no choice but to bite the bullet here and concede that the view is unable to accommodate this intuition; we cannot rule out the possibility of perverse peak experiences, such as Genghis Khan's. If a peak experience causes seemingly perverse changes in an agent's values, there is nowhere to stand *outside* of any individual agent's web of moral and normative judgments from which to say that that particular agent's values have *worsened* as a result of those changes. The constructivist cannot say, *Yes, now that you hold the perverse values X, Y, and Z, your subjective motivational set is a lot worse than it used to be*. Or, to be more precise, the constructivist *can* say this, but implicit in his saying this is that he is making the judgment with reference to his own evaluative standpoint (or someone else's – or perhaps some combination of particular evaluative standpoints). In other words, one agent is passing judgment on the changes in another agent's evaluative standpoint and is doing so with tacit reference to the standards that are contained in his own, particular evaluative standpoint.

As Street (2010: 366) puts it, 'there are no standards of correctness in the normative domain [which encompasses the moral domain] except ... from the point of view of someone who already accepts some normative judgments or other – the point of view of a valuing creature'. To return to the example of Genghis Khan, if Genghis Khan forms a perverse new evaluation during a peak experience – e.g. 'It is better to be absolutely merciless in dealing with the inhabitants of city-states that oppose us because then the reputation of the Mongols will grow more fearsome over time, and in the long run, it will mean that foreign cities are more likely to surrender to us immediately than attempt to resist, which will reduce the need for us engage in protracted sieges of fortified cities' – there is no *stance-independent* place from

which we can say that this new addition to Genghis Khan's motivational set is an 'objective' change for the worse (it just seems like a change for the worse from the point of view of our own evaluative standpoints).

Of course, the constructivist can still criticise Genghis Khan, in a genuinely stance-independent way, for failures of coherence or a failure to appreciate the relevant non-normative facts. For instance, perhaps Genghis Khan's new value is implicitly predicated on factual falsehoods or mistaken assumptions. If it is, then *this* fact obtains independently of the evaluations of any particular agent. The facts are the facts, no matter what any individual agent's evaluations are. But the constructivist must concede there are no 'objective', stance-independent standards that we can appeal to in order to judge whether a particular substantive change in an agent's values represents a step in the right direction or in the wrong direction; we can only render such judgments with tacit reference to our own evaluative standpoints.[30]

The best that the enlightened constructivist can do with respect to this objection, the perverse peak experiences objection, is to engage in a form of damage control. That is, the enlightened constructivist can try to show that perverse peak experiences are, in reality, vanishingly rare and that peak experiences almost always change an agent's values in a way that seems, *both from that agent's own point of view and from the point of view of other agents*, like a significant, higher-order-stable improvement (though, once again, there is no stance-independent way of establishing that this corresponds to an *actual* improvement). Unfortunately, I do not think that enlightened constructivism, as it currently stands, has the resources to do even this much. As we saw from the Genghis Khan example earlier, it is relatively easy – too easy – to think of plausible examples of perverse peak experiences. The main reason for this is that the emotion of very intense joy, which chiefly characterises peak experiences, can be had in so many contexts that seem pathological. We can imagine, for instance, a torturer during the Spanish Inquisition

---

[30] For what it is worth, I personally do not find this result to be at all unintuitive, but I can see how, from the evaluative standpoint of other agents, it might seem unintuitive.

having a peak experience while torturing a heretic in a dungeon[31]; or a suicide bomber having a peak experience moments before detonating a bomb in a crowded place, all the while believing that he is doing God's work; or a professional jewellery thief – like from the movie *The Pink Panther* – having a peak experience after having pulled off the biggest jewellery heist of his life, etc.

What I want to do in the next section, then, is to explore whether we can modify the enlightened constructivist method so as to narrow the range of possible perverse peak experiences that one could have under the view, thereby rendering their occurrence vastly less likely in the actual world. I do not think that there is any way for the enlightened constructivist to rule out the bare possibility of an agent's having moral insights during an authentic peak experience that seem utterly perverse and deranged from the perspective of other agents. By comparison, and as discussed in Chapter 3, §3.3, there is no way for the (Humean) constructivist to rule out the bare possibility of hypothetical amoralists, agents who are stipulated to be ideally coherent, informed, imaginative, etc., but who are perfectly indifferent to and unmoved by moral considerations. But what the enlightened constructivist can do is to try and show that the bare possibility of perverse peak experiences is not as alarming or revisionist as it might, at first blush, seem – in other words, to show that the revisionist element of the theory is tolerably rather than intolerably revisionist.

I hasten to add that my reasons for introducing this modification to enlightened constructivism are not just *ad hoc*. I am not trying to modify enlightened constructivism merely in pursuit of extensional adequacy, merely to try and circumvent the perverse peak experiences objection. Rather, I have independent reasons for exploring this modification, which I will unpack in the next section.

---

[31] Lest this seems like a somewhat caricatured and unrealistic example, consider the following quote from Pinker (2012: 157): 'Torture in the Middle Ages was not hidden, denied, or euphemized … No, torture was woven into the fabric of public life. It was a form of punishment that was cultivated and celebrated, an outlet for artistic and technological creativity. Many of the instruments of torture were beautifully crafted and ornamented. They were designed to inflict not just physical pain, as would a beating, but visceral horrors, such as penetrating sensitive orifices, violating the bodily envelope, displaying the victim in humiliating postures, or putting them in positions where their own flagging stamina would increase their pain and lead to disfigurement or death. Torturers were the era's foremost experts in anatomy and physiology, using their knowledge to maximize agony, avoid nerve damage that might deaden the pain, and prolong consciousness for as long as possible before death.'

4.5 WHITHER ENLIGHTENED CONSTRUCTIVISM? THE RELEVANCE OF MEDITATION PRACTICE TO THE CONSTRUCTIVIST METHOD

I have been dropping hints throughout this chapter about an interesting modification to the enlightened constructivist method. Let me now elaborate. I want to explore whether we can tweak the constructivist method such that instead of drawing on peak experiences in the construction of ethical truth, we draw on *meditative* peak experiences: peak experiences that are induced through meditation practice.

I have already discussed in this chapter how peak experiences have the phenomenology of seeming profoundly morally and normatively insightful. As Aaltola (2015: 289) puts it, peak experiences often have the power to 'spark foundational change in one's moral outlook'. I think that this fact provides us with an independent reason for trying to incorporate them into our theory of value, *whatever that theory turns out to be in the end* (not just constructivism). I want to explore whether certain types of meditative experience share this phenomenology of seeming profoundly morally insightful – I am not claiming that they all do – and, if so, whether they might be accommodated within the theory of value that is constructivism.

I will sketch out this experimental meditative constructivist view in §4.7 and §4.8. Before getting to that, however, the first port of call is to define meditation practice. I will devote a significant amount of time to this definitional task; it will take up the whole of §4.6. The reason for this is twofold. First, as will become clear later on, there are many different forms of meditation practice, and it is a considerable task in itself to identify which forms of meditation might be relevant to meta-ethical theorising – I am not suggesting that they all are.

Second, as I mentioned in the introduction to this thesis, the phenomenology of meditation is a topic where I feel that I am able to make an original contribution to the existing philosophical literature. Insofar as one of the key goals of this thesis is to try and make some novel and interesting contributions to ongoing philosophical conversations, I think my discussion of meditation practice is a worthwhile digression.

I will argue in §4.7 that there is one particular of type of meditative experience that does have the phenomenology of seeming, from the first-person point of view of the agent undergoing it, to be ethically insightful. That is the state of intense,

boundless love, kindness, and compassion that is elicited and cultivated through loving-kindness meditation. I go into detail about what exactly is involved in the practice towards the end of §4.6.

For now, let me just say that loving-kindness meditation is inherently other-regarding in nature. The *explicit goal* of the practice is to elicit a state in which an 'unconditional feeling of loving-kindness and compassion pervades the whole mind as a way of being, with no other consideration, or discursive thoughts' (Lutz et al. 2008: 1). Bankard (2015: 2336–2340) offers the following useful take on the ethical salience of loving-kindness meditation:

> Current neuroscientific evidence strongly suggests that moral emotions such as compassion [and love and kindness] can be honed using various forms of Eastern meditation … By cultivating a twinge of compassion, one increases the likelihood of compassionate behaviour. Thus, the habit of meditation cultivates prosocial behaviour by honing [certain moral emotions] … Short-term [loving-kindness] meditation training can lead to an increase in prosocial behaviour during a novel task completely unrelated to the training itself … LKM training increases prosocial behaviour toward people who were not specifically targeted during the training … LKM training can have lasting effects on compassionate behaviour.

Given the pronounced other-regarding dimension to the practice, the state of mind that is the object of the practice is of particular interest here. In §4.8 I will sketch out an experimental constructivist view that incorporates this state into the method for constructing truths about one's moral reasons. Ultimately, I will argue that this experimental view is too beset with problems to really be viable. But despite this pessimistic conclusion, I hope to show the view is an interesting and quirky position in logical space, and is worth exploring.

4.6 WHAT EXACTLY *IS* MEDITATION?

The task of defining meditation practice is not an easy one. I think the best approach here is oblique rather than direct: rather than baldly stating what meditation is, I will begin by offering a bit of context. I will then slowly creep up on a definition. I should

mention that the definition of meditation I present is novel and original insofar as it does not draw on any previous definitions of meditation that I have encountered; I think that my definition does a better job of distilling the essence of the practice, as far as I understand it.

Over the past two decades, many studies have found that there are considerable health benefits to regular meditation practice. Meditation reduces anxiety, depression, and chronic pain (Goyal et al. 2014: 357). Meditation boosts emotional intelligence and relieves perceived stress (Chu 2010: 177). Meditation bolsters immune function, improves sleep quality, and lowers blood pressure (Rose et al. 2020: 508). Meditation precipitates structural and functional changes in the brain, such as an increase in grey matter density in regions associated with learning and memory, emotional regulation, and self-awareness (Harris 2014: 35–36; Krygier et al. 2013: 306). There is now a vast literature on the physiological and mental effects of meditation practice (e.g. Zeidan et al. 2010; Zeidan et al. 2011; Hölzel et al. 2011; Kim et al. 2010; Godfrin and Van Heeringen 2010).

But despite all of the attention that meditation has garnered in recent years, there is a relative dearth of academic work on the first-person, experiential, phenomenological side of the practice (Przyrembel and Singer 2018: 82–83). Almost all academic articles concerning meditation provide only a very brief, cursory summary of what is involved in the practice at the level of direct, first-person experience – often just one or two lines. The following summary offered by Rose et al. (2020: 507) is typical in this regard: 'Meditation is defined as a form of cognitive training that aims to improve attentional and emotional self-regulation … Meditation that involves fixation on a specific mantra, image, feeling, or idea is known as focused attention; meditation that involves fixation on the present moment is known as mindfulness.' I am not suggesting that this definition is inaccurate, but it is so cursory as to be potentially misleading to any reader who does not already have some prior familiarity with meditation practice.

Given how much scholarly interest there is in the third-person, objective effects of meditation, I think it is important to have in place a clear account of the first-person, experiential, subjective side of the practice. In laying out my definition of meditation practice, I hope to provide just such an account. I aim to do so in a way that is: (a) jargon-free; (b) secular (in the sense of being free of any controversial, Buddhistic

metaphysical suppositions); (c) accessible to those who have no prior familiarity with or interest in meditation practice. As mentioned in §4.5, my main reason for introducing the topic of meditation is that I think it is relevant to the issue of modifying the enlightened constructivist method. However, my secondary reason for introducing the topic is that I hope to provide a novel way of understanding what is involved, experientially, in the practice of meditation. It is in this sense that I hope to make an original contribution to the ongoing philosophical conversation regarding meditation practice (and its connection to peak experiences).

Let me begin by noting some key features of our everyday, *non*-meditative experience. The default waking state of the human mind is to be habitually and almost continuously engaged in internal self-talk. We live our lives with a kind of running interior monologue. This internalised self-chatter includes 'visual images, odd observations, snatches of melody, fun facts, old grudges, pleasant fantasies, memorable moments' (Pinker 2015: 79). From earliest childhood to decrepit old age, our experience of reality comes filtered through the lens of discursive thought. By 'discursive thought', I mean the ideas, opinions, images, bits of language, etc., which continuously arise in the mind and which tend to pass from topic to topic in a rambling, digressive way. For instance, in the privacy of our own minds we sometimes rehearse past conversations: going over what we said, what we didn't say, what we ought to have said. We fantasise and agonise about the future, producing an endless stream of imagined scenarios that fill us with hope or dread. We tell ourselves the story of the present as though there were a blind man inside our heads who required continuous narration to know what is happening (Harris 2014: 94).

Subjectively, it feels almost like we are listening to the radio inside our own minds. We are tuned into an internal radio station that never stops broadcasting. This internal radio channel forms a kind of background hum to our waking lives. I should be clear that the term 'thought', as I am using it, encompasses a wide variety of mental appearances, such as ideas, opinions, observations, notions, reflections, resolutions, memories, daydreams, judgments, etc. It does not, however, include perceptual experiences, such as the sound of a hive of buzzing bees nearby. Let me add as well that I am taking the terms 'internal self-chatter' and 'running interior

monologue' to be synonymous with 'discursive thought' (for the purposes of this chapter).

So, my claim is that we spend much of our waking lives engaged in internal self-talk and that we do so without knowing that we are doing so (i.e. we are lost in discursive thought without realising it). What makes this point quite difficult to communicate is that *when* thoughts arise in the mind, we can't help but feel somehow identified with them: we feel like our thoughts are *us*, in some sense. It is almost as though our thoughts are a facet or feature of ourselves, somewhat analogous to ripples along the surface of a body of water: something distinct from the water yet inseparable from it. But there is a bit of a paradox here – again, I am talking about our everyday, *non*-meditative experience for now – because although we implicitly feel identified with our thoughts, there is also an implicit sense that we are somehow *causing* our thoughts. Generally speaking, the sense is that we have an active role in *thinking* our thoughts, as opposed to just being the passive recipient of them. I hasten to add that there are exceptions to this. People suffering from paranoid schizophrenia, for instance, might have an experience of certain thoughts *intruding* into their consciousness, like unwelcome visitors coming into a house without having been invited.

There is a further layer of complexity here that makes things even harder to parse. We often have emotional/evaluative/judgmental reactions in response to our thoughts. That is, we have higher-order thoughts: thoughts about thoughts. Indeed, there is a kind of open-ended recursivity to human thought (Pinker 2012: 224). Thoughts can bend back on themselves indefinitely, like an infinity mirror. Imagine, for instance, that one is engaged in some activity that one normally finds enjoyable, such as going for a brisk walk on a crisp Autumn day. Imagine that in the course of this activity, a painful, unpleasant memory bubbles to the surface of awareness. This could be the memory of some professional setback, such as botching an important public-speaking engagement. When unpleasant memories like this arise, they exert a kind of gravitational pull over the mind, warping one's attention and polluting one's mood. The memory of a botched, subpar public-speaking performance, say, may cause one to start replaying and litigating the event in one's mind: going over what one said, what one did not say, what one ought to have said. The arising of an unpleasant memory – which, after all, is just a thought – can set off a chain reaction

of concatenating negative thoughts. And one's attention can become captivated by these concatenating thoughts for some time. For instance, one might spend five minutes so engrossed in the memory of a negative past event that one entirely forgets that one is engaged in some activity that is meant to be enjoyable, like going for a walk on a crisp Autumn day.

The proneness of the human mind to engage in internalised self-chatter places a demand on one's attention: it uses up cognitive bandwidth (Schooler et al. 2011: 319). I am here operating with William James' well-known definition of attention: '[Attention] is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others' (James 1890: 322). Internalised self-talk diverts one's attention *away* from the information streaming in through the various sensory channels and *towards* whatever new idea, opinion, observation, etc., is arising in consciousness.[32] The technical name for this phenomenon is 'perceptual decoupling' (Schooler et al. 2011: 319). This is when attention becomes mostly disengaged from perception and instead focuses on 'mental events that arise without any external precedent (known as stimulus independent thoughts)' (ibid.).

Broadly speaking, meditation is a form of systematic mental training that allows one to silence all internal self-chatter, even if only for seconds at a time. In meditation, one gradually strengthens one's ability to self-regulate one's attention. Rather than multi-tasking in one's own head and habitually dividing one's attention between the inner stream of discursive thought and the facets of perception, one silences all internal rumination and devotes one's full and undivided attention to the percepts in one's experience. There is an experiential alternative to feeling identified with the next thought that arises in the mind. Meditation allows one to access this experiential alternative. One can punctuate one's day with moments of clear-seeing, during which time one's attention is very lucid and crystalline and perception-focused (at least, this is how things *seem* when meditating; I will revisit this caveat further on in this section). This meditative clarity is not a matter of achieving a 'new conceptual understanding or adopting new beliefs about the nature of reality'; meditation is not

---

[32] I will define the term 'consciousness', which I am taking to be synonymous with the term 'awareness', further on in this section; the definition will be relevant to some of the arguments that I make later on.

about getting *more information* into one's head or *better quality information* (Harris 2014: 38). Rather, it is a matter of experiencing the present moment 'prior to the arising of thought' (ibid.).

At this point, many might object that experience is *itself* a type of thought, and so, talk of 'experience unmediated by thought' is nonsensical. Suffice it to say that my claim is more precise than that articulated in Harris' quote above. My claim is that there is a systematic methodology for training one's attention – i.e. meditation – that allows one to experience reality unmediated by all internal self-chatter (what we might call '*discursive* thought', as opposed to 'thought' in its broadest sense). Furthermore, the felt, first-person sense of being the *thinker* of one's thoughts, of somehow *causing* one's thoughts to arise in awareness, is an illusion. It is an illusion that can be collapsed, experientially, in a way that is conclusive.[33]

This point about experiencing the world unmediated by internal self-chatter may still seem a bit vague, so let me try and concretise it by taking a slightly different tack. If one pays clear, close, and undisturbed attention to any facet of one's experience, that facet can start to become uncanny and unfamiliar. Take, for instance, a vision of colour – the colour green, say. If one can silence one's internal self-talk for long enough to undistractedly observe a particular instance of the colour green as it *actually is*, the fact that one is in total ignorance of *what it is* can become plain. What actually *is* the colour green? Not as a function of wavelengths of light or neurophysiology but as it is directly perceived. There really is nothing to say but that it is 'green'. In fact, it is not even 'green', which is just a word. The vision of colour that we see before us just *is* what it ineffably is in each moment. The moment we suspend the conceptual associations that we have with a given object or percept – i.e. our knowledge of truths *about it* – our direct experience *of it* can shade into this experience of dumb mystery. What we are left with, at that point, is a non-lingual, non-conceptual intuition regarding certain contents of consciousness, about which nothing more can really be said.

Before I proceed, let me just acknowledge that it is highly philosophically controversial whether an experience can ever be truly non-conceptual in the sense of being entirely unmediated by concepts of any kind. Broadly speaking, concepts are mental representations that commonly take the form of mental images or word-

---

[33] I will argue for this claim further on in this section.

like entities (Margolis and Laurence 2021: 3–4). They occur as part of propositional attitudes, and what they represent 'contributes to the content of these attitudes, to what we believe, desire, etc.' (Glock 2009: 7). Concepts are generally conveyed to others by means of language, gestures, drawings, etc. (Jackendoff 1995: 22).

In characterising meditative experiences as 'non-conceptual', I am not suggesting that there are *no* necessary conceptual requirements on experience. Rather, I am making the weaker claim that meditative experiences generally involve stripping away many – but not all – of the concepts that one operates with during periods of mundane waking consciousness. For example, meditative experiences usually involve effacing those concepts that manifest in the form of mental images or word-like entities (i.e. those concepts that form part of what I have been describing as one's 'running internal monologue'). However, during a meditative experience, one might still be tacitly employing certain very general or demonstrative concepts (e.g. one might still have some vague sense of one's position in space and in relation to one's surroundings).

To return to the point at hand, no philosopher has better captured this experience – the experience of everyday objects suddenly becoming profoundly uncanny and strange – than Jean-Paul Sartre. The protagonist of Sartre's first novel, *Nausea*, is a brooding young historian by the name of Antoine Roquentin. At one point in the novel, Roquentin has a deeply unnerving encounter with a tree:

> So I was in the park just now. The roots of the chestnut tree were sunk in the ground just under my bench. I couldn't remember it was a root anymore. The words had vanished and with them the significance of things, their methods of use, and the feeble points of reference which men have traced on their surface. I was sitting, stooping forward, head bowed, alone in front of this black, knotty mass, entirely beastly, which frightened me. Then I had this vision.
>
> It left me breathless. Never, until these last few days, had I understood the meaning of "existence." I was like the others, like the ones walking along the seashore, all dressed in their spring finery. I said, like them, "The ocean is green; that white speck up there is a seagull," but I didn't feel that it existed or that the seagull was an "existing seagull"; usually existence hides itself. It is

there, around us, in us, it is us, you can't say two words without mentioning it, but you can never touch it. When I believed I was thinking about it, I must believe that I was thinking nothing, my head was empty, or there was just one word in my head, the word "to be." Or else I was thinking . . . how can I explain it? I was thinking of belonging, I was telling myself that the sea belonged to the class of green objects, or that the green was a part of the quality of the sea. Even when I looked at things, I was miles from dreaming that they existed: they looked like scenery to me. I picked them up in my hands, they served me as tools, I foresaw their resistance. But that all happened on the surface. If anyone had asked me what existence was, I would have answered, in good faith, that it was nothing, simply an empty form which was added to external things without changing anything in their nature. And then all of a sudden, there it was, clear as day: existence had suddenly unveiled itself. It had lost the harmless look of an abstract category: it was the very paste of things, this root was kneaded into existence. Or rather the root, the park gates, the bench, the sparse grass, all that had vanished: the diversity of things, their individuality, were only an appearance, a veneer. This veneer had melted, leaving soft, monstrous masses, all in disorder—naked, in a frightful, obscene nakedness (Sartre 2000: 182–183).

There are different ways of interpreting this passage. I think one plausible construal is that Roquentin has stumbled into a quasi-meditative experience. As he sits there on the park bench, looking down at the black tangled mass of the root, he is suddenly shunted into experiencing the world as a raw, immediate, amorphous, Heraclitean flow of phenomena. The familiar, comfortable concepts of everyday life have melted away. Roquentin's internal self-chatter has died down to a whisper. He has been left with an unsettling vision of reality unmediated by discursive thought.

I mentioned earlier that, broadly speaking, meditation is a form of systematic mental training that allows one to silence one's internal self-chatter and gradually strengthen one's ability to self-regulate one's attention. Let me now get more precise about what exactly is involved in the practice. In trying to get more precise, however, one runs into the immediate obstacle that 'meditation' is really an umbrella term, somewhat

analogous to the term 'sports'. There are dozens of different kinds of practices that are grouped under the heading of 'meditation', just as there are dozens of different kinds of physical activities that are classified as 'sports'.

I do not have the space to exhaustively taxonomise all of the different strands of meditative practice. Instead, I will focus on the three most influential and prominent forms: focused-attention meditation, mindfulness meditation, and loving-kindness meditation. Once I have finished defining these three forms of meditation, I will explore whether the rarefied experiences that are induced through these three methodologies are at all similar to *peak* experiences, and if so, whether they have an ethical dimension to them in the sense outlined in §4.2 (and if so, whether this provides us with a way of modifying the enlightened constructivist method along the lines set out in §4.5).

Before proceeding, let me add a quick note regarding a claim I made in the prefatory remarks to this section. I mentioned that the definition of meditation I present is original. In making this claim, I was referring specifically to *mindfulness* meditation. However, I think that my construal of focused-attention meditation also adds something new to the existing literature on this topic. My definition of loving-kindness meditation is, by contrast, much more deferential to existing scholarship. The reason for this is that the existing definitions of loving-kindness meditation do, in my view, an adequate job of capturing the essence of the experiential side of the practice.

Let us begin with focused-attention meditation. By far the most well-known variant of focused-attention meditation is *vipassanā*, so I will focus on that. *Vipassanā* is the Pali word for 'insight' (Harris 2014: 34). *Vipassanā* comes from the oldest tradition of Buddhism, the Theravada (ibid.). *Vipassanā* typically directs the meditator to focus their attention on the in-and-out cycle of breathing (Ricard, Lutz, and Davidson 2014: 41). *Vipassanā* can be thought of as the cognitive practice of: (a) sustaining attention on the variegated sensations of the breath; (b) monitoring discursive events – i.e. episodes of mind-wandering – as and when they occur; (c) disengaging from those discursive events without affective reaction; (d) re-focusing attention back on the breath (Zeidan et al. 2011: 5540).

This summary is quite abstract. Let me make things more concrete by distilling the *vipassanā* technique into a series of direct instructions. I should note that these instructions reflect my own interpretation of *vipassanā*, though this interpretation is, I believe, convergent with that of others who have written extensively on this topic (e.g. Goldstein 2013: 45–54; Harris 2014: 39–40).

1. Find a secluded spot where you can practice meditation undisturbed. The Buddha suggests going to 'the forest or to the root of a tree or to an empty hut' (Ñāṇamoli and Bodhi 1995: 145). But a quiet room or office space should suffice.
2. Take a seat, keeping your spine erect.
3. Close your eyes, inhale deeply a few times, and slowly become aware of the various points of contact between your body and the chair that you are sitting in. Notice the sensations associated with sitting: feelings of warmth, pressure, tingling, etc.
4. Gradually become aware of the process of breathing. Try to focus your attention wherever you feel the breath most distinctly – either in your nostrils or in the rising and falling of your chest or solar plexus.
5. Do not attempt to control the breath; just let it come and go naturally. Pay attention to the sensation of breathing as if it were the only thing of any consequence, the only thing worth attending to. Pay attention as if you were Adam or Eve, newly created in the Garden of Eden, experiencing the sensation for the very first time. Do not let any of the raw sensory data go unobserved; do not lose any of the experience.
6. As you meditate on the breath, various distractions will arise, such as feelings of impatience and restlessness. Let these distractions pass by. Keep your attention trained on the breath. It may be helpful to think in terms of the following metaphor. You are rowing a boat down a river. You are trying to focus purely on the act of rowing: on the feel of the oars in your hands and the rhythmic motion of your muscles. Unfortunately, there are people standing on either bank of the river who keep calling out to you, trying to get your attention and distract you from your task. These

people represent the unwelcome distractions that you will face when trying to meditate. Ignore them, and keep on rowing.

7. Every time you notice your mind wandering in thought – e.g. thinking about all of the tasks that you have to complete later that day – gently return your attention to the breath.

Hasenkamp et al. (2012) have used brain imaging to identify the neural networks activated by focused-attention meditation. They have shown that there is a four-phase cognitive cycle involved in focused-attention meditation: 'an episode of mind wandering, a moment of becoming aware of the distraction, a phase of reorienting attention, and a resumption of focused attention' (Ricard, Lutz, and Davidson 2014: 42).

So, with focused-attention meditation, one begins by focusing one's attention on a specific subset of the contents of consciousness, such as the sensations associated with the breath. I say that one 'begins' this way because as one proceeds further with the practice, the meditation can expand to encompass other facets of one's experience. For instance, once one builds up sufficient concentrative momentum in one's *vipassanā* practice, one might, in addition to paying attention to the breath, also start noticing any subtle undercurrents of mood or emotion that are colouring one's experience. One might notice a background feeling-tone of anxious impatience or a sense of doubt and futility regarding the practice. The dispassionate observance of all these emotional subtexts can become part of the meditative practice.

Let me flesh out this point about background moods and emotions colouring one's experience by way of an analogy. When sailing a boat out at sea, one can often perceive a gust of wind coming long before it actually hits the boat: one can see the gust coming from miles off as it ruffles the surface of the water. Out at sea, large gusts of wind look almost *embodied*; they seem somehow more tangible and perceptible than they do on land. When on land, one tends not to notice a gust of wind until one is in the midst of it and feels it blowing all around one. Likewise, if one is *not* being mindful of moods and emotions, then one tends to only notice them once they are blowing through the gates and alleys of the mind.

Just as a keen-eyed sailor can spot a south-westerly wind long before it actually hits the ship, so, too, the keen meditator can spot subtle emotions and moods percolating on the periphery of awareness long before they build up enough momentum to rock the boat of the mind. What is more, being mindful of a mood like irritability, say, can: (a) reduce its half-life to an impressive degree; (b) sever the link between that mood and whatever behavioural imperatives it seems to communicate. Typically, when one is in an irritable mood, that mood affects the way that one interacts with other people and the way that one performs various tasks. A task that one might otherwise have found enjoyable, such as doing some gardening, can become a frustrating experience. However, if one pays clear, careful, dispassionate, undistracted attention to whatever negative mood is present, that very quality of attention can serve to *denude* the mood, to strip it of whatever power it once had to derange one's life. Irritability, when clearly and undistractedly observed, is really nothing more than a feeling of jittery energy coursing through the body (this should be read as an experiential claim rather than a metaphysical one).

The quality of mind that one cultivates during *vipassanā* practice is known as 'mindfulness'. Mindfulness is a state of 'clear, nonjudgmental, and undistracted attention to the contents of consciousness, whether pleasant or unpleasant'; it is a 'vivid awareness of whatever is appearing in one's mind or body – thoughts, sensations, moods – without grasping at the pleasant or recoiling from the unpleasant' (Harris 2014: 35–36). Mindfulness is a meta-cognitive state of awareness (Szekeres and Wertheim 2014: 374). That is to say, it is a mental state that arises when attention is explicitly directed towards noting the current contents of consciousness (Smallwood and Schooler 2015: 495). Mindfulness is characterised by a perspectival shift regarding internal sensations, emotions, thoughts, etc. (Szekeres and Wertheim 2014: 374). Instead of reacting to these phenomena and trying to change them, one merely *witnesses* them with the detached calm of a dispassionate observer; one witnesses these phenomena with an almost scientific interest, as if one were trying to discern their experiential signature as carefully as possible. In short, focused-attention meditation – i.e. *vipassanā* – is the practice, and mindfulness is the state of mind generated by the practice. It is a state of mind that many find to be incredibly lucid, equanimous, and tranquil (Harari 2011: 442).

There is a great deal to unpack here, specifically on the question of whether *mindfulness* is best characterised as a peak experience and, if so, whether it has an ethical dimension to it. But before I get to that, let me first finish defining the other two key forms of meditation practice: mindfulness meditation and loving-kindness meditation. Given that I have just spoken about how mindfulness is the state of mind cultivated during focused-attention meditation, it may seem confusing that there is another form of meditation with the word 'mindfulness' in the title. So, what exactly is mindfulness meditation, and how does it differ from focused-attention meditation?

Unlike focused-attention meditation, mindfulness meditation does *not* involve initially focusing one's attention on a specific subset of the contents of consciousness (e.g. the sensations associated with the breath). Rather, the idea is to pay clear, unclouded, indiscriminate attention to whatever sensation, emotion, etc., is arising in the present moment (Ricard, Lutz, and Davidson 2014: 42). This involves taking an open, non-judgmental stance towards all phenomena that surface in awareness. It involves witnessing these phenomena unfolding in a continual procession without allowing one's attention to become distracted by discursive thought. One must cease thinking about what has happened in the past or what is going to happen in the future. One must instead pay clear, careful attention to what is happening in one's experience right now, in the present moment. Indeed, *as a matter of direct, first-person experience*, the past does not even exist. The past is nothing but a series of thoughts arising in the present moment (e.g. vivid and evocative memories). Likewise, the future is a mere figment of the imagination, a kind of waking dream occurring in the present. When the future arrives, it will simply *be* the present moment. So, subjectively speaking, we inhabit a kind of timeless, eternal present. In mindfulness meditation, one is orienting one's attention *away* from the inner flow of discursive thought and *towards* the raw, unmediated facets of one's experience in the present moment.

So, as one sits there practising mindfulness meditation, one might notice, for instance, the early signs of a headache coming on: a kind of dull pulsing in the forehead. And this may, in turn, prompt a small surge of annoyance. And this surge of annoyance may trigger an unpleasant memory of a less-than-optimal interaction that one recently had with a family member or work colleague. The aim is to dispassionately observe all of these goings-on without becoming entangled in them,

to observe them with a meta-cognitive level of awareness (i.e. mindfulness), almost as if they were happening to someone else. One can observe whatever feeling of annoyance is colouring one's experience, say, without having this feeling imbue one's subjectivity in such a way as to make one feel identified *with it*. It is as though one were a scientist trying to study all of the phenomena that comprise one's first-person experience as carefully and dispassionately as possible.

The key difference between focused-attention meditation and mindfulness meditation lies in the initial scope of the practice. Focused-attention meditation – e.g. *vipassanā* – directs the practitioner to intentionally beam the spotlight of their attention at a limited subset of the contents of consciousness (e.g. the sensations associated with the breath). The idea is that by deliberately narrowing the focus of one's attention in this way, one can gradually hone the ability to self-regulate one's attention. Once this ability is strengthened to a sufficient degree, one can then start incorporating other facets of one's experience into the practice (e.g. bodily sensations). In theory, if one walks the path of focused-attention meditation for long enough, one reaches a point where the practice encompasses all aspects of one's experience in the present moment. It is at this point that the practice morphs into a maximally expansive form of mindfulness meditation. The two practices, focused-attention meditation and mindfulness meditation, become one.

The following analogy may help to clarify things. We can think of the project of cultivating meditative ability as being analogous to the project of climbing a high mountain with the goal of enjoying the sublime, expansive view from the summit. There are two paths up the mountain. One path is gradual and winds around and around the mountain like a coil, slowly leading one up to the summit. The other path is an incredibly steep, straight line going directly from the base of the mountain to the summit at the top. One can walk the former path without straining oneself too hard; this path has a gentle gradient but is also so long and circuitous that many people never reach the end of it. The latter path, on the other hand, represents a far more direct and expeditious route to the summit. The problem with this direct path is that it is near-vertical at certain points, which makes for arduous climbing. In this analogy, the long, gentle path stands for focused-attention meditation and the steep, direct path stands for mindfulness meditation. In theory, both paths lead to the same place:

the summit. The summit stands for a state of mind of pristine, untrammelled mindfulness: maximally-mindful mindfulness, if you will.[34]

In reality, the difference between focused-attention meditation and mindfulness meditation is a bit more subtle and complex than the above remarks suggest. In order to fully appreciate the difference, we need to revisit a claim I made earlier (I made this claim with a promissory note to argue for it in more detail further on; the time has now come to cash that promissory note). The claim is that meditation allows one to experientially collapse the felt, first-person sense of being the *thinker* of one's thoughts. To state it baldly, the *self* is an illusion. This claim is absolutely critical to understanding the difference between focused-attention meditation and mindfulness meditation, so let me now explain and argue for it.

The term 'self', as I am using it, refers to the felt, first-person sense of being a locus of awareness inside the head that can strategically direct attention at certain appearances in consciousness (Harris 2014: 126). This notion of 'self' is described by Olson (1998: 648) as 'Account 2' of the self. It is the idea that one's self is the inner subject of one's conscious experiences – a subjective centre of gravity, as it were, that exists somehow interior to the body. Let me phrase things in more direct language. One feels like one is located *behind one's eyes*. One feels like one is *behind one's eyes* in a way that one is not *behind one's knees*, say. One is 'up here', and one's knees are 'down there'. One feels like one is *located in space*: like one is inside one's head looking out at the world. One does not feel synonymous with one's body. Rather, one feels like the owner or inhabitant of one's body: like a locus of awareness that exists somehow interior to the body: like an inner subject that can use the body as a kind of object. Another way of coming at this is to imagine oneself in a different body (Garfield 2022: 8). Imagine waking up one morning in the body of King Charles III, say. It is *this* feeling, the feeling that one is a subject existing interior to the body that could hypothetically be transplanted into a different body, that I am referring to when I talk of the 'self'.

One can take this feeling of self as the object of meditation. One can look for the one that is looking, for the subject, for the 'I', and this very act of looking can cause that feeling of self to drop out of conscious experience altogether. This is compelling

---

[34] I will unpack what this state consists of further on in this section.

in the same way that the disappearance of any illusion is: 'You thought something was there, but upon closer inspection, you see that it isn't. What doesn't survive scrutiny cannot be real' (Harris 2014: 116). In various schools of Buddhism, this notion of the illusoriness of the self is referred to as *ānatta* (Struhl 2020: 114).

To make things a bit more precise, let me distil the foregoing remarks into the following argument:

(1) One of the key features of everyday conscious experience is for one to feel – at least implicitly – that one is a locus of awareness existing interior to the body that is able to strategically direct attention at particular appearances in consciousness (e.g. sounds coming from one's local environment).

(2) The term 'self' is often used to refer to this feeling (Olson 1998: 648; Struhl 2020: 114).

(3) If one subjects this feeling of self to sustained introspective scrutiny – i.e. if one *turns attention back on itself*, back on this feeling of being located behind one's face, inside one's head – that very act of introspective scrutiny will reliably cause the feeling of self to drop out of one's experience altogether.

(4) Features of one's experience that do not survive sustained and careful introspective scrutiny of this kind cannot be real.

Therefore,

(5) The self, as defined above, is not a veridical feature of one's experience: it is an 'illusion that functions at the phenomenal level' (Struhl 2020: 115).

Let me address one obvious objection to this argument. The sense of being the inner subject of one's conscious experience cannot be an illusion because in order for it to be *experienced* as an illusion, there must be some subject of experience that is *doing* the experiencing; experience cannot be without a subject. But this objection elides a distinction we might draw between two different notions of a subject. There is a *false* subject, which is just the feeling one has of being a locus of awareness inside the head, behind the face. This is the notion of a subject that I am taking the term 'self' to refer to. But then there is the *true* subject of experience, which is just the prior condition, the prior context, of open, undefined, mirror-like awareness *in*

which the feeling of self is arising and being known. This true subject of experience is what I have been referring to as 'awareness' or 'consciousness' thus far.

Everything that one takes as first-person evidence for the existence of a self – i.e. the false subject – is actually arising in a broader landscape of mind. This broader landscape of mind is consciousness/awareness (I am taking the two terms to be synonymous for the purposes of this chapter). Consciousness is 'prior to [the feeling of self], a mere witness of it, and, therefore, free of it in principle' (Harris 2014: 104). *As a matter of direct, first-person experience*, one is truly identical to consciousness, as opposed to any of the contents of consciousness – which includes the feeling that there is a centre *to* consciousness. Consciousness is mode-neutral yet has 'intrinsic phenomenal character' (Albahari 2009: 63). By 'mode-neutral', I mean that consciousness is not restricted to any particular sensory or cognitive modality, such as seeing, hearing, smelling, tasting, touching, and thinking.

To be clear, the claims I am making regarding consciousness are *experiential* rather than *metaphysical*. I am not presenting a metaphysical, third-person, objective account of what consciousness is. I am merely describing how things *seem*, as a matter of direct, first-person experience, when one when proceeds far enough with the practice of meditation. With this caveat on board, let me elaborate on the notion of consciousness that I am appealing to here.

Consciousness is like the reflectivity of a mirror. It is like a mirror with no surface and no frame, a disembodied mirror. Every facet of one's experience appears within this space of pure reflectivity; every facet is reflected in this disembodied mirror. And yet, this mirror-like space is intrinsically uncontaminated by anything that appears within it. Again, the analogy with a normal mirror is apropos. A mirror can reflect a raging fire so completely that it seems indistinguishable from that fire, and yet, when the fire is extinguished, we see that the mirror was always something distinct from the flames that it so vividly reflected. Every high and low of human experience – whether it be intense euphoria, serene tranquillity, existential angst, etc. – is an appearance arising in consciousness, the disembodied mirror. And that which is *aware* of intense euphoria, or whatever else, is not *itself* intensely euphoric. Consciousness is unable to observe itself directly. Nonetheless, it is aware of itself as something immediately present in experience (Albahari 2009: 65–66). It cannot be known in the same way that we know objects appearing *within* it, such as the

sight of a large alligator, the feeling of warm sunlight on one's face, or the smell of freshly-ground coffee. Consciousness is the 'quality of bare knowing of whatever is seen, heard, felt, or cognized' (Goldstein 2013: 42–43). It is a kind of substratum to all experience.

With the above in mind, what mindfulness meditation *truly* consists of is as follows. Mindfulness meditation is the practice of cancelling oneself down subjectively until one is nothing but an open, undefined, mirror-like space of awareness. (Again, this is a purely experiential claim; I am not attempting to do metaphysics here.) One then rests *as* this space of awareness and merely witnesses whatever arises within it. This could include sounds from one's local environment, patterns of light, colour, and shadow in one's visual field, sensations of pressure, tingling, and temperature in one's body, feelings of frustration, impatience, or irritation with the practice of meditation itself, etc. The aim of the game is to discern with pristine clarity what is subjectively real in each moment. When one is successfully practising mindfulness meditation, it is as though one has brought oneself up against the very edge or lip of experience, to its *event horizon*, and is watching sights, sounds, sensations, etc., unfold in real time. It is as though one has dispelled layers of distraction that had hitherto been hanging like vapours before the eye and can now behold the irreducible, elemental core of one's experience in the present moment.

The state of mind that one cultivates during mindfulness meditation is very similar to the state of mind that one cultivates during focused-attention meditation. In both cases, it is the state of *mindfulness*. Recall that I defined mindfulness earlier as a meta-cognitive state of awareness that arises when attention is explicitly directed towards noting the current contents of consciousness and which is characterised by a perspectival shift regarding internal sensations, emotions, thoughts, etc. (Szekeres and Wertheim 2014: 374; Smallwood and Schooler 2015: 495). The key difference between the state of mindfulness that is induced through focused-attention meditation and that which is induced through mindfulness meditation is that the latter generally involves collapsing the illusion of self in a direct way. That is to say, it takes aim at the very feeling that one is explicitly directing attention towards noting the current contents of consciousness: it takes this feeling of being able to direct attention as *just another content of consciousness*.

Let me return to my earlier analogy of there being two paths up the same mountain. Focused-attention meditation represents the gradual, circuitous path up the mountain. Mindfulness meditation represents the steep, direct path. Both paths lead to the same place: the summit. The summit represents a state of pristine, untrammelled mindfulness wherein the meditator dispassionately observes any and all objects of awareness as and when they arise – e.g. sights, sounds, sensations, intentions, reactions, judgments, perceptions, ideas, opinions, observations, notions, reflections, resolutions, memories, etc. Crucially, these objects of awareness *include* the feeling of being a self that is able to direct the flow of attention at other objects of awareness. In this pristine, untrammelled state, one has cancelled oneself down subjectively until one is merely the open, undefined, mirror-like condition in which every aspect of one's experience is arising and then passing away in a ceaseless Heraclitean parade of phenomena. I will hereafter refer to this state of mind as 'perfect mindfulness'.

One problem with this analogy is that it gives the misleading impression that the state one attains at the summit of the mountain is somehow permanent or fixed. In reality, the state of perfect mindfulness is something that one can slide in and out of across time. One might sustain this mental state for 100 seconds, say, before then lapsing into distraction to a greater or lesser degree (e.g. becoming lost in discursive thought). But the more one hones one's ability to self-regulate one's attention through meditation, the more frequently one can access this state and the longer one can sustain it. Focused-attention meditation and mindfulness meditation are both systematic methods for accessing and cultivating the state of perfect mindfulness. However, they have different ways of going about it.

All versions of focused-attention meditation are, at least to begin with, implicitly predicated on the illusion of self. Let us take *vipassanā* practice as an example (*vipassanā* being by far the most well-known variant of focused-attention meditation). As discussed earlier in this section, *vipassanā* involves paying clear, careful, undistracted attention to the sensations associated with breathing (Ricard, Lutz, and Davidson 2014). This kind of meditation is implicitly dualistic: when one pays attention to the breath in this way, the implicit feeling is that one is located *above* the breath; one feels like one is located *inside one's head* and is aiming one's attention *downwards* at the breath. The logic of *vipassanā* practice is to gradually build up

one's ability to self-regulate one's attention. Once this ability reaches a certain point, one can then begin incorporating other facets of one's experience into the practice, such as bodily sensations. In theory, if one proceeds far enough with the practice of *vipassanā*, one eventually has the astonishing realisation that there *is* no self who is self-regulating one's attention (Goldstein 2013). In other words, once the ability to self-regulate one's attention is honed to a sufficient degree, something quite strange happens, which is that attention can be turned back on itself: one can look for the one that is looking, for the subject, for the 'I', for the *source* of attention, and can see that there *is* nothing to find. Any feelings associated with attention or the honing of one's attention are themselves just appearances in consciousness, appearances that must be recognised *qua* appearances if one is to access a state of perfect mindfulness.

One issue with the *vipassanā* approach is that the implicitly goal-oriented nature of the practice can prevent one from recognising that consciousness is *already* intrinsically free of anything that could be called a 'self' (Bodian 2017; Harris 2014). The implicit logic of *vipassanā* practice is that one must build up sufficient concentrative momentum in order to induce an experience of perfect mindfulness, mindfulness free of self. One is trying to produce an experiential effect based on meditative effort. The implicit view is that this kind of self-transcending mindfulness is not coincident with ordinary waking consciousness. One cannot experience perfect mindfulness while taking one's poodle for a walk, say.

So, practitioners of *vipassanā* are implicitly paying attention to the breath, bodily sensations, undercurrents of mood, etc., *in order to* induce some rarefied, transcendent state of mind (or many such experiences). But the strange truth is that the very state of mind these practitioners are seeking to induce – or, at least, that many of them are ultimately seeking to induce – is *intrinsic to the nature of awareness*. This is because the state of perfect mindfulness *just is* what it feels like when one cancels oneself down subjectively until one is merely the open, undefined, mirror-like condition, the prior context, in which every aspect of one's experience in the present moment is arising and then passing away. It is almost as though the state of mind that one is seeking is *too close to be seen*: it is right before one's eyes, and yet it is obscured by the habitual and deeply-ingrained tendency of the human mind to engage in near-constant internal self-talk and to contract itself into a feeling

of *self*, a locus of awareness inside the head that is implicitly presumed to be at the centre of this ever-blooming efflorescence of sights, sounds, sensations, and so on.

Mindfulness meditation, by contrast, generally points one in the direction of this experiential truth in a more direct way. As an example, let us take the form of mindfulness meditation that is practised in the Dzogchen branch of Tibetan Buddhism (pronounced 'zo-chen') (Harris 2014: 132). In Dzogchen meditation, it is taught that the state of mind of perfect mindfulness can be accessed at any moment because consciousness is *already* free of self, already self-transcending. One is not trying to produce an experiential effect based on meditative effort. One is not trying to induce a transcendent, rarefied state of mind through some heroic act of concentration. Rather, one must recognise how the mind *truly is* in each moment (as a matter of experience). The very state one seeks is intrinsic to the nature of awareness: it just is what it feels like to drop back and 'rest' as the bare fact of awareness, the borderless, mirror-like condition in which various appearances are ceaselessly arising and then passing away, like clouds in the sky. And it is this condition that one is truly experientially identical to, not anything arising within it (which includes the feeling of being a meditator *doing* the meditation, of trying to effortfully elicit some transcendent state of mind).

Before I move on to defining the final major form of meditation practice, which is loving-kindness meditation, let me add one final word on the difference between mindfulness meditation and focused-attention meditation. In the preceding paragraphs I may have given the impression that mindfulness meditation is somehow superior to focused-attention meditation. But in truth, both approaches have their upsides and downsides. I have already spoken about one of the downsides of focused-attention meditation, which is that the implicitly goal-oriented nature of the practice – of trying to produce an experiential effect based on meditative effort – can prevent one from accessing the very state that one is seeking to access: perfect mindfulness. But in the interest of being even-handed, let me acknowledge that there are also some significant downsides to mindfulness meditation, as typified in the Dzogchen tradition.

For one thing, it can be difficult to communicate to people what the practice actually consists of. The nature of the practice is not as easily comprehensible as in

versions of focused-attention meditation, such as *vipassanā*. The reason for this is that in Dzogchen meditation – which I am taking to be representative of mindfulness meditation more generally – there is far less emphasis than in *vipassanā* on gradually and progressively building up one's ability to self-regulate one's attention. One is thrown straight into the deep end of the swimming pool, as it were. Quite early on in Dzogchen practice, one is instructed to cut through the illusion of the self and to drop back and 'rest' as the bare fact of awareness, the borderless, mirror-like condition in which every facet of one's experience is arising and being known (Harris 2014: 132). That is to say, one is put on a steep and direct path to the state of perfect mindfulness. To those who are taking up meditation for the first time, this instruction may seem inscrutable, nonsensical, or indicative of some kind of category error. By contrast, when one takes up *vipassanā* practice for the first time, one is taught to begin by paying clear, careful, undistracted attention to the sensations associated with breathing (Ricard, Lutz, and Davidson 2014). This instruction is relatively easy to understand and put into practice. And once one has done so, one can gradually build up from there, following the systematic methodology laid out in *vipassanā* teachings.

We can draw an analogy with learning to ride a bike. *Vipassanā* practice is a little bit like learning to ride a bike with the training wheels on. If one is riding a bike with the training wheels on, then one is not *really* riding a bike. One is just performing an activity that looks a lot like riding a bike, the purpose of which is to develop one's motor control, balance, etc., to the point where one is capable of truly riding a bike. (This is not to say that the activity couldn't still be enjoyable and rewarding, but just that one is not really doing it properly yet.) Dzogchen meditation, by contrast, is like learning to ride a bike with no training wheels on – and perhaps with no handlebars, either. If one can make it work – i.e. if one can push the pedals and get the wheels to turn, all the while keeping one's balance – then one can be said to be authentically riding a bike right from the start. One has leapfrogged over the training period. But how many people who have never ridden a bike before are able to keep their balance on their very first attempt? Not many. And if one falls during the attempt and breaks one's leg, then that may put one off riding a bike for life.

Let us now turn to the final major form of meditation practice, which is loving-kindness meditation. Loving-kindness meditation 'cultivates attitudes and feelings of loving kindness and compassion toward other people, whether they are close relatives, strangers or enemies' (Ricard, Lutz, and Davidson 2014: 43–44). The practice involves attuning oneself to someone else's needs and then experiencing a 'sincere, compassionate desire to help that person or to alleviate the suffering of other people by shielding them from their own destructive behaviour' (ibid.). Bankard (2015: 2336, citing Lutz et al. 2008: 1) offers the following useful summary:

> Typically, [loving-kindness meditation] involves quiet contemplation in a seated position. Attention is then directed toward warm and tender feelings for a particular person (e.g. a spouse, parent, or good friend). Slowly these warm emotions are expanded to include the person meditating, the meditator's community, humanity as a unified whole, and eventually the world. The goal of such meditation is to produce a state in which an "unconditional feeling of loving-kindness and compassion pervades the whole mind as a way of being, with no other consideration, or discursive thoughts."

Loving-kindness meditation often involves the silent repetition of a mantra conveying benevolent intent, such as *May you be happy; may you be free from suffering; may your life unfold with ease* (Stell and Farsides 2016: 141). The repetition of a mantra can act as a focal point for one's feelings of love, kindness, and compassion, allowing them to gradually augment with each repetition until they reach a fever pitch of intensity and seem to pervade one's whole being. The expectation is that once one proceeds far enough with the practice, one will be able to direct these feelings of love, kindness, and compassion even towards those whom one would normally find odious or repulsive. Kearney et al. (2014: 33) offer some further details on this point:

> When practising LKM, a person is asked to notice and feel the positive emotion elicited by each of the phrases of positive intention, or to notice whether there is a sense of reluctance, hesitation, or even aversion for the self or another [e.g. a neighbour whom one finds very irritating and obnoxious]. The practitioner is encouraged to greet these responses with

kindness and to notice them without judgment. More generally, during LKM, when practitioners become distracted by thoughts that arise during the practice, they are instructed to notice what has distracted them with an attitude of nonjudgmental, mindful attention, and then return to the LKM phrases and the breath without self-criticism.

One interesting aspect of loving-kindness meditation is that it aims to elicit feelings of intense *compassion* rather than feelings of intense *empathy*. Empathy can be understood as 'the act of coming to experience the world as you think someone else does' (Bloom 2018: 16). If a well-to-do businesswoman is walking along the sidewalk on a cold winter evening, say, and sees a homeless man huddled under some old blankets, shivering from the cold, and she feels vicarious suffering in response to his real, physical suffering, then what she is experiencing is empathy. Empathy involves mirroring other people's feelings in this way. Compassion, on the other hand, does not. Compassion is an emotional as well as motivational state that one experiences when one is concerned for 'another's suffering and desires to enhance that person's welfare' (Leiberg et al. 2011: 1). It is characterised by feelings of 'warmth, love, and concern for the other as well as the desire to help and promote the other's welfare' (ibid.).

To return to the example of the well-to-do businesswoman, let us say that after experiencing an initial empathetic twinge of vicarious suffering in response to the suffering of the homeless man, she resolves to take positive action to improve the situation. She ducks into a nearby café and buys a cup of hot soup and a bagel, which she then gives to the man, hoping to lift his spirits a little. The homeless man thanks her effusively, and she goes on her way, basking in the warm halo of having done a good deed. If the woman goes about this action – of buying the homeless man a cup of hot soup and a bagel – with energy and good cheer, if she *feels good* while doing good, then this emotional and motivational state is best characterised as one of compassion rather than empathy. She began by feeling the man's pain in an empathetic way, and that feeling was painful for her. But she then took the positive step of doing something to ease the man's plight, and in the course of performing that action, she felt warmth, love, and concern, not suffering.

Research has shown that compassion is associated with positive emotions, while empathy, by contrast, can lead to 'empathy fatigue', wherein the empathiser reaches a point of emotional exhaustion or burnout (Ricard, Lutz, and Davidson 2014: 44). To vicariously feel the suffering of others can be draining. Compassion does not drain one in this way; instead, it 'reinforces an inner balance, strength of mind, and a courageous determination to help those who suffer' (ibid.). In support of this claim, loving-kindness meditation has been shown to benefit 'health care workers, teachers and others who run the risk of emotional burnout linked to the distress experienced from a deeply empathetic reaction to another person's plight' (ibid.).

Before moving on to the next section, I think it is worth clarifying one further aspect of loving-kindness meditation. I have spent a lot of time in this section talking about the illusoriness of the self, especially as it relates to the differences between focused-attention meditation and mindfulness meditation. Given this, one might wonder whether loving-kindness meditation also involves collapsing this feeling of self. The answer is that loving-kindness meditation is comparable to focused-attention meditation in that it is, to begin with, implicitly predicated on a kind of experiential dualism: one feels like a locus of awareness inside the head that is *performing* the meditation, that is *directing* one's attention towards 'warm and tender feelings for a particular person', say (Bankard 2015: 2336). However, just as with focused-attention meditation, if one proceeds far enough with the practice, one eventually dissolves this feeling of self. If one succeeds in eliciting feelings of transcendent love, kindness, and compassion through loving-kindness meditation, these feelings may reach such a zenith of intensity that they (temporarily) obliterate the sense of being a self, of being an inner subject hovering somewhere behind one's eyes, of *performing* the meditative practice; it is as though one has been swept away by a tidal wave of love, kindness, and compassion and has been so thoroughly tumbled in the swell as to have had the structure of one's subjectivity rearranged – it is like one has been turned inside out by the force of these extraordinary emotions.

I should say that my reason for raising the topic of the self in the first place was merely to clarify the (quite subtle) differences between focused-attention meditation and mindfulness meditation. And I only raise the topic again here, in my discussion of loving-kindness meditation, lest its omission seems like an oversight. But this point about the illusoriness of the self has no further role to play in my argument; the next

section, as its title suggests, focuses on the question of whether meditative experiences have an ethical dimension to them (in the sense outlined in §4.2), and the topic of the self and its claimed illusoriness is not relevant to that question, and hence will not come up again in this chapter.

4.7 ARE MEDITATIVE EXPERIENCES MORALLY INSIGHTFUL?

So, we have seen what is involved, as a matter of direct, first-person experience, in the three most influential and prominent strands of meditation practice: focused-attention meditation, mindfulness meditation, and loving-kindness meditation. Let us now turn to the question of whether the states of mind that these three methodologies aim at cultivating are at all similar to *peak* experiences and, if so, whether they are ethically insightful.

As discussed in the preceding section, both focused-attention meditation and mindfulness meditation aim to cultivate the state of mind of perfect mindfulness. Just as a reminder, what is involved in this state is the dispassionate observance of any and all objects of awareness – e.g. sights, sounds, sensations, intentions, reactions, judgments, perceptions, ideas, opinions, observations, notions, reflections, resolutions, memories, etc. Crucially, these objects of awareness *include* the feeling of being a self that is able to direct the flow of attention at other objects in awareness. In a state of perfect mindfulness, one has cancelled oneself down subjectively until one is just the open, undefined, mirror-like condition, the prior context, in which every aspect of one's experience is arising and then passing away in an ever-flowing, Heraclitean stream of phenomena (again, this is a purely experiential claim, not a metaphysical one).

Perfect mindfulness is one kind of meditative experience. The other kind – at least, the other kind that is pertinent to our discussion – is the state of mind of extraordinary love, kindness, and compassion that is the object of loving-kindness meditation. For shorthand, I will refer to this state as 'ethereal love'. I will argue in this section that perfect mindfulness is *not* a peak experience and does *not* have an ethical dimension to it, but that ethereal love *is* a peak experience and *does* have an ethical dimension to it. Let us begin with the former claim.

As discussed in §4.6, there is something paradoxical – or seemingly paradoxical – about the state of perfect mindfulness. Technically, it is a mental state. But it is

misleading to think of it as a mental state that can be elicited in the same way as other mental states, such as euphoric love or expansive compassion. Perfect mindfulness is not an experiential *effect* that one is trying to produce based on meditative effort. It is not as though one needs to sit there, straining every nerve in order to build up sufficient concentrative momentum to unlock an experience of perfect mindfulness. The state is intrinsic to the nature of awareness and can be accessed at any moment (though, as covered in §4.6, this is easier said than done).

The key point is this. In cultivating the state of perfect mindfulness, one is not trying to create a fireworks display of extraordinary positive emotions. Any powerful emotions that arise either during focused-attention meditation or mindfulness meditation are a by-product of the process rather than its ultimate object (for the sake of simplicity, I will focus on mindfulness meditation and put focused-attention meditation to one side for now). If one is to be perfectly mindful, one must dispassionately observe whatever emotions arise in the mind without becoming identified with them. If one starts feeling blissfully happy while meditating, then one must observe this feeling in the same way that one observes bodily sensations, like a pain in one's lower back. Indeed, there is a danger in any extraordinary emotions that one encounters while walking the path of mindfulness meditation because such states can distract one from the true goal of the practice, which is accessing and cultivating the state of perfect mindfulness. Such states can lead one to draw the erroneous conclusion that mindfulness meditation is about eliciting *peak experiences*, which it is not. Any peak experience that one happens to have while practising mindfulness meditation (or focused-attention meditation) is just another appearance in consciousness – albeit a very vivid and captivating one. It is an appearance in that prior condition of open, undefined, mirror-like awareness that one is truly identical to (as a matter of experience).

So, not only is the meditative experience of perfect mindfulness *not* a peak experience, it is orthogonal to all peak experiences. During a peak experience, one is swept away on a tide of incredible joy, say. During perfect mindfulness, by contrast, one merely observes such emotions – and whatever other experiential phenomena go towards constituting a peak experience – as and when they arise. One observes them with pristine, detached clarity, as a scientist might observe through a microscope the movements of microscopic organisms in a petri dish. The

whole point is *not* to get caught up in emotions of this kind or in any other appearances in awareness, such as ideas, memories, daydreams, etc.

Whatever peak experience one succeeds in eliciting through meditation practice will, by its very nature as an evanescent phenomenon, also pass away. As the Buddha observed, 'Whatever has the nature to arise will also pass away' (Goldstein 2013: 284–285). The transient mental states that are involved in peak experiences are very different from the unchanging mental state that characterises perfect mindfulness. I say 'unchanging' because the state is intrinsic to the nature of awareness; we might think of it as the base note or feeling-tone of awareness itself. The state is always there to be accessed, and it does not vary each time one accesses it. The state just is what it feels like to 'drop back' and rest as the bare fact of awareness, the open, undefined, mirror-like condition, the prior context, in which every facet of one's experience is arising unbidden and then passing away, like clouds in an expanse of sky.

We have established, then, that perfect mindfulness is not a peak experience. But it could still be relevant to modifying the enlightened constructivist method if it had an ethical dimension to it (in the sense outlined in §4.2). So, does it have an ethical dimension to it? The answer is 'no'. This is because any apparent moral insights, revelations, epiphanies, etc., that arise when one is in a state of perfect mindfulness are just more appearances in awareness that one must dispassionately observe if one is to *maintain* that state of perfect mindfulness. As soon as one becomes preoccupied with a seeming moral epiphany and starts *thinking* about it, turning it over in one's mind, considering its implications, etc., then one has lapsed into a form of distraction and has degraded the quality of one's mindfulness. One's mindfulness has slipped from being perfect to imperfect because one's attention has become captured by what are, in fact, just transient appearances in consciousness.

So, for example, let us say that one is sitting there, practising mindfulness meditation and that one accesses the state of perfect mindfulness. Imagine that one maintains this state for a period of 120 seconds but that the state is then interrupted by the arising of a constellation of thoughts that distract one's attention. This constellation of thoughts strikes one as being a profound moral realisation. We can imagine these thoughts taking the following form. *This is all one long, unfolding dream. The people you meet are characters in a dream. The objects that come into*

your possession are dream-objects. As you walk, as you converse, as you do anything, you can almost hear the sand trickling through the hourglass, the threads of the tapestry unwinding. And when we arrive at the terminus of the dream, the final few moments before we tip into oblivion, we will surely regret our failures of kindness, our failures to make the dream as morally beautiful as it could have been. Each act is a sacrament – or can be.

The details do not matter so much here. The point I am making is that if a cluster of thoughts arises in one's mind in the course of practising mindfulness meditation, and if those thoughts strike one as being a profound moral realisation, then that realisation is *orthogonal* to the state of perfect mindfulness. It represents a declension in the quality of one's mindfulness. One is no longer dispassionately observing the objects of awareness, merely witnessing them, but is becoming captivated by certain of those objects.

To return to an analogy that I used in §4.6, when one is in a state of perfect mindfulness it is almost as though one has brought oneself to the very edge or lip of experience, to its *event horizon*, and is watching sights, sounds, sensations, etc., unfolding in real time. If any moral insights happen to arise when one is in that state, then one must observe them as raw experiential phenomena. If one's attention becomes captivated by them – if one starts mulling them over, considering their implications, etc. – then it is almost as though one has taken a step back from the edge: one is no longer standing on the event horizon of one's experience in the present moment but is merely in its general vicinity. The key takeaway here is that the state of mind of perfect mindfulness does *not* have an ethical dimension to it, and as such, it is not relevant to the enlightened constructivist method for constructing truths about one's moral reasons.

Let us now consider the state of mind of ethereal love that is elicited through loving-kindness meditation. I should note that not much actually hinges on whether this state qualifies as a *peak* experience or not; this is a purely taxonomical point. The real mission here is to determine whether the state has an ethical dimension to it and, if so, whether we might be able to modify the enlightened constructivist method such that instead of drawing on the evaluations that are revealed or elevated during

conventional peak experiences to construct truths about one's moral reasons, we draw on those that are revealed or elevated during experiences of ethereal love.

Let me begin by briefly restating the definition of peak experiences that I presented in §4.1. Peak experiences are experiences of very intense joy – often combined with very intense love – that are short in duration and which involve a feeling of heightened clarity, insight, and higher-order stability on the part of the individual experiencing them; they generally have an ethical/normative dimension to them in that they allow one to see the possibility of a more coherent version of one's set of values and/or prompt transformative changes in one's values; they are perspective-altering in that they grant one a seemingly more profound perspective on life, one which is rich with meaning and significance; they are sufficiently rare and extraordinary that they stand out in stark contrast to the experiences that surround them in time and space. As mentioned in §4.1, this is a stipulative definition aimed at clarifying and elucidating the principal subject matter of this chapter: peak experiences. I am taking the conditions stated in the definition to be jointly sufficient. None of the conditions, however, are individually necessary (I discussed the reason for this back in §4.1, which is that the subject matter we are dealing with here – i.e. peak experiences – is inherently nebulous and ineffable, and I want to allow some flexibility in my definition so that an experience which diverged only very slightly from the stated conditions might still qualify as a peak experience).

The state of ethereal love that is elicited through loving-kindness meditation shares almost all of the features mentioned in the above definition but diverges in a few minor particulars. The first is that while peak experiences are primarily characterised in terms of the emotion of joy, experiences of ethereal love are primarily characterised in terms of the emotion of love. To be more precise, ethereal love is characterised by a synaesthetic blend of three key emotions: love, kindness, and compassion. These emotions come together like musical notes to form a single, harmonious chord. This is not to say that the state of ethereal love could not also involve the emotion of joy. The experiences one has during loving-kindness meditation can be joyful, certainly. But ethereal love is not *primarily* characterised in terms of joy. Likewise, I am not claiming that peak experiences could not involve the emotions of love, kindness, and compassion. Indeed, I mention in my stipulative

definition that peak experiences often involve a feeling of intense love. I am merely suggesting that peak experiences are not primarily characterised by these emotions.

The second minor point of divergence relates to duration. Peak experiences tend to be short in duration, on the order of a couple of minutes in length (see §4.1 for more discussion of this point). The state of ethereal love can be, though is not always, longer in duration. It is hard to quantify the average duration of ethereal love as, to my knowledge, meditators in this tradition do not attempt to record the duration of the state on those occasions when they succeed in eliciting it. But given that loving-kindness meditation is a systematic methodology that is explicitly geared towards cultivating this particular state of mind, one would expect experiences of ethereal love to be, on average, both longer in duration than peak experiences and also more frequent. For what it is worth, my own personal – albeit limited – experience with loving-kindness meditation seems to corroborate this point: when one focuses one's attention in a systematic way on cultivating the state of ethereal love, experiences of that state, when they do occur, tend to be both longer-lasting and more frequent than your average peak experience.

This point about duration carries implications for another condition stated in my definition: peak experiences are sufficiently rare and extraordinary that they stand out in stark contrast to the experiences that surround them in time and space. This condition may not be met by experiences of ethereal love. Imagine that one sits down and practises loving-kindness meditation for one hour and that during that time, one succeeds in inducing a state of ethereal love on three separate occasions, each lasting for several minutes. Given that one has experienced ethereal love three times in a single hour, it seems a bit tenuous to say that any one of those experiences stands out in 'stark contrast' to the experiences that surround it in time and space – though I suppose it depends on what we take to be the relevant context here: if that single hour of loving-kindness meditation is the only hour in which one has experienced ethereal love in the course of a whole month, say, then one's experiences of ethereal love may still satisfy the condition stated above.

These are minor quibbles, minor points of divergence. On the whole, experiences of ethereal love have almost complete overlap with peak experiences. As such, I think that they are best thought of as a particular kind of peak experience. To draw a zoological analogy, we can think of experiences of ethereal love and peak

experiences as being a bit like two isolated populations of the same species. Let us take lions as an example. There are two main kinds of wild lions existing in the world today: Asiatic lions and African lions. The vast majority of wild lions are African lions, and they live south of the Sahara (Britannica Academic 2023: 'lion'). A small minority of wild lions are Asiatic lions; there are approximately 650 of them living in an isolated population in India (ibid.). In this analogy, the larger population of African lions represents conventional peak experiences, and the smaller population of Asiatic lions represents experiences of ethereal love. They are both lions; they are both the same species. But there are minor points of divergence between them (e.g. Asiatic lions have a shorter and sparser mane). Anyway, the key takeaway is that the state of mind of ethereal love counts as a peak experience (for the purposes of this thesis, at least).

Let us now turn to the more important question. Does ethereal love have an ethical dimension to it, and if so, is it relevant to modifying the enlightened constructivist method for generating truths about one's moral reasons? Let me begin by quoting a useful summary from Bankard (2015: 2336–2340) of the ethical salience of loving-kindness meditation:

> Current neuroscientific evidence strongly suggests that moral emotions such as compassion [and love and kindness] can be honed using various forms of Eastern meditation … By cultivating a twinge of compassion, one increases the likelihood of compassionate behaviour. Thus, the habit of meditation cultivates prosocial behaviour by honing [certain moral emotions] … Short-term [loving-kindness] meditation training can lead to an increase in prosocial behaviour during a novel task completely unrelated to the training itself … LKM training increases prosocial behaviour toward people who were not specifically targeted during the training … LKM training can have lasting effects on compassionate behaviour.

The above extract introduces the term 'moral emotion'; moral emotions are those that are 'linked to the interests or welfare either of society as a whole or at least of persons other than the judge or agent' (Haidt 2003: 853). Given that the emotions of

love, kindness, and compassion that characterise the state of ethereal love all have a strong other-regarding dimension to them, they qualify as *moral* emotions under Haidt's definition, which is the definition that I will operate with throughout this chapter.

Bankard (2015: 2340) draws a connection between the moral emotions that are cultivated through loving-kindness meditation and pro-social, other-regarding behaviour; as he mentions in the quoted matter above, training in loving-kindness meditation can increase such behaviour during a 'novel task completely unrelated to the training itself' and towards agents who were 'not specifically targeted during the training', and these effects are 'lasting' rather than temporary. This is an interesting and commendable effect of loving-kindness meditation. However, it is not directly relevant to the enlightened constructivist method. Recall from the introduction to this chapter that the enlightened constructivist is interested in bolstering or 'shoring up' the strength of our moral *reasons*, not in increasing our pro-social behaviour. So, the key question is whether there is a mechanism by which the moral emotions that are honed via loving-kindness meditation *could* result in a strengthening of our moral reasons (I am not yet claiming that they do, in fact, result in such a strengthening; that part of the argument will come later).

The answer to this question is 'yes'. This is because moral emotions 'impact moral judgments' and 'often lead to [new] moral judgments' (Bankard 2015: 2328). Recall that under constructivism, moral judgments are evaluations, which are conative states that provide their possessor with a goal and a motivation to pursue that goal. Evaluations differ from ordinary beliefs in that they are intrinsically motivational. The moral emotions of love, kindness, and compassion that characterise the state of ethereal love generally prompt a reconfiguration of one's evaluations. This could involve: (a) causing the formation of new evaluations; (b) increasing the strength of certain existing evaluations; (c) decreasing the strength of certain existing evaluations; (d) eliminating certain evaluations entirely; (e) some combination of (a) through (d).

Under a constructivist view of things, the strength of a moral reason co-varies with the strength of the relevant evaluations in an agent's subjective motivational set. (By 'relevant', I just mean whatever evaluations go towards determining truths about a particular moral reason that one possesses.) The moral emotions that are involved

in loving-kindness meditation could alter one's relevant evaluations – e.g. one's pro-social, cooperative, altruistic, other-regarding, etc., states of valuing – in such a way as to generally strengthen one's moral reasons. The caveat 'generally' is important here because, as discussed at length in Chapter 3, §3.8, the relevant non-normative facts have a crucial role to play in determining what moral reasons one possesses.

4.8 EXPERIMENTS IN CONSTRUCTIVISM: INTRODUCING 'MEDITATIVE CONSTRUCTIVISM'

When one succeeds, through loving-kindness meditation, in augmenting the moral emotions of love, kindness, and compassion until they reach a fever pitch of intensity, when one's whole being seems infused and radiant with these emotions, the feeling is one of heightened ethical insight, of seeing ethical matters in a clearer light, of having discovered a more other-regarding and compassionate way of being in the world, one which feels like a vast improvement over the status quo.

Indeed, when one comes down from this experience, an experience of ethereal love, the feeling is almost of being shrunken and contracted back into one's everyday self, like a genie that is being forced back into its bottle. It can be depressing to notice the distance between how loving, kind, and compassionate one could be, if only one could live from a place of ethereal love all the time, and how loving, kind, and compassionate one tends to be during periods of mundane waking consciousness.

The practice of loving-kindness meditation is explicitly and inherently other-regarding in nature. Recall from my summary of the practice in §4.6 that one begins by directing one's attention towards 'warm and tender feelings for a particular person (e.g. a spouse, parent, or good friend)' (Bankard 2015: 2336). One kindles these warm emotions – almost like when one is kindling a fire and cups one's hands around a small, fledgling flame to prevent the wind from blowing it out – until they slowly expand outwards to encompass 'the meditator's community, humanity as a unified whole, and eventually the world' (ibid.). The explicit goal of the practice is to elicit a state – what I'm calling 'ethereal love' – in which an 'unconditional feeling of loving-kindness and compassion pervades the whole mind as a way of being, with no other consideration, or discursive thoughts' (Lutz et al. 2008: 1).

Given the essentially other-regarding nature of loving-kindness meditation, and given that moral emotions play a more prominent role in experiences of ethereal love

than in conventional peak experiences, I think it is true to say that experiences of ethereal love generally involve a *greater* subjective feeling of ethical insight and illumination than regular, non-meditative peak experiences.

To be clear, I am not suggesting that conventional peak experiences do not also involve moral emotions; as discussed in §4.1, they sometimes involve feelings of intense love. But the state of ethereal love that is elicited through loving-kindness meditation is different in that the core of the experience consists of the three moral emotions of love, kindness, and compassion. These three emotions are not a possible off-shoot of the experience, a possible direction that it could take, but are right at the centre of it, like the molten core at the centre of a planet.

I argued at length in §4.2 that the felt sense of heightened moral insight during a conventional, non-meditative peak experience corresponds to seeing the possibility of a more coherent version of one's set of values *and/or* undergoing transformative changes in one's values that are higher-order stable. We can apply this same interpretation to experiences of ethereal love. The moral insight one gains during an experience of ethereal love might involve, for instance, seeing for the first time how certain values are in tension with one another, grasping the full implications of certain values, becoming suddenly cognisant of values that had hitherto been implicit or buried in dark recesses of the mind, etc. It might also involve substantive changes in one's subjective motivational set that go beyond just remedying failures of coherence, such as increasing the strength of those evaluations that are pro-social, cooperative, altruistic, and other-regarding in character.

Let me offer an example to make all this a bit more concrete. Imagine that we have two brothers: Heinrich and Matthias. Heinrich and Matthias are estranged from each other: they haven't spoken in years. As teenagers, they were always intensely competitive – always competing over who was the better athlete, the better student, the more doted on by their parents, etc. – and this sibling rivalry continued into adulthood. As they grew older, the brothers' life paths diverged. Heinrich, the elder of the two, ended up having a successful and lucrative career as a management consultant. He also ended up marrying his childhood sweetheart and having a happy marriage, the product of which was three healthy and loving children. Matthias, the younger brother, ended up becoming a history teacher at a secondary school. He is

twice divorced and has no children. Let us say that the brothers had a bitter falling out at a family reunion. During the reunion, Matthias felt that Heinrich was subtly mocking him for his comparative lack of professional and financial success and trying to embarrass him in front of the rest of the family. A heated argument broke out between the brothers, which resulted in their coming to blows; Matthias dumped a bowl of potato salad on Heinrich's head, and Heinrich hurled a tray of canapés at Matthias. Since then, the brothers have shunned each other's company.

Now imagine that Matthias, the younger brother, decides to give loving-kindness meditation a try, acting on a recommendation from a friend. Matthias practises diligently, spending many hours in 'quiet contemplation in a seated posture', slowly kindling 'warm and tender feelings in an open-hearted way' (Fredrickson et al. 2008: 1046). At a certain point in the practice, he follows the meditative instructions of directing these feelings at someone with whom he has a troubled and fractious relationship, which, in his case, is his brother Heinrich. When he first tries this, there is a 'sense of reluctance, hesitation … [and] even aversion' (Kearney et al. 2014: 33). But he follows the meditative instructions of 'greet[ing] these responses with kindness and [trying] to notice them without judgment' (ibid.). He keeps at it, and eventually he succeeds in eliciting a state in which an 'unconditional feeling of loving-kindness and compassion pervades the whole mind as a way of being, with no other consideration, or discursive thoughts' (Lutz et al. 2008: 1). That is to say, he elicits an experience of ethereal love.

Imagine that during this experience, Matthias is flooded by love, kindness, and compassion. He feels that he has never experienced such intense love in his entire life. It is as though he is walking above the clouds, bathed in the exquisite light of the upper atmosphere, breathing the rarefied air. And from that lofty position he feels that he is seeing things more clearly than ever before. In a sudden flash of insight he realises that his antagonism towards his older brother is, and always has been, rooted in envy. It is as though a floodlight is illuminating the dark recesses of his mind, revealing evaluations that he had held implicitly up till that point but which he had never really been conscious of. He sees now that he has always tacitly judged himself to be more worthy and deserving of success in life than his brother; he has judged himself to be a deeper person than his brother, with a less shallow, less trivial character; he has always resented his parents for subtly favouring his brother over

him, as though his brother were the golden child and he just a pale imitation; he thinks it is profoundly unfair that fortune has left him twice divorced and childless, while his brother has been blessed and blessed again, etc.

As Matthias sees all of this, the layers of buried resentment that have accumulated over the years, he realises how petty and childish it all is and what an opportunity cost it represents. Indeed, he is repulsed by it. He now views this aspect of his psychology, his deeply-rooted jealousy and wounded pride, as something warping and malignant, like a tumour that needs to be excised. He resolves to wipe the slate clean. He has not spoken to his brother, his only sibling, in five years. He sees now what a moral failing it was to have let this childish feud drag on for so long. He ought to have swallowed his pride, reached out, and reconciled with his brother long ago. He decides to call his brother on the phone that very day and arrange a time to meet. It is the first step on the path to rebuilding their relationship.

This little tale of Matthias and Heinrich is just a vignette; I am trying to give a sense of the ethical dimension of experiences of ethereal love, to give a sense of the sort of ethical insights that one might have when in this state. I think I have said enough on this point, so let me now move on to discussing some independent motives for exploring a meditation-based modification to constructivism.

Peak experiences tend to be haphazard and unpredictable: they strike one out of the blue. One might be hiking solo through the Lake District, say, and as one crests the top of a ridge and glimpses the panoramic views beyond, one might be suddenly flooded by a feeling of seraphic joy. Such experiences, as I have argued at length in this chapter, can be morally insightful and rich in profundity, but they are also infrequent and unsystematic – one stumbles into them as if by accident.

My tentative suggestion here is that meditation practice could systematise the having of certain kinds of peak experiences. Instead of stumbling into peak experiences more or less at random, one could use this methodology for training one's attention to elicit peak experiences in a way that is more reliable and replicable. I argued in the previous section that the state of perfect mindfulness that is cultivated through both focused-attention meditation and mindfulness meditation *does not count as a peak experience and does not have an ethical dimension to it*. So, when I talk about meditation being a methodology for reliably eliciting certain

kinds of peak experiences, *I am referring to the methodology of loving-kindness meditation and experiences of ethereal love*.

Towards the end of §4.6 I discussed what the practice of loving-kindness meditation consists in, so I won't repeat what has already been said. Suffice it to say that loving-kindness meditation is a systematic, step-by-step guide for eliciting and sustaining states of ethereal love. If one spends five hours hiking solo on the Isle of Skye, say, one *may* have a peak experience during that time, but there is no guarantee of one's doing so – and it would not at all be surprising if one did *not* do so.

By contrast, if one spends five hours diligently practising loving-kindness meditation, it *would* be surprising if one did not, at some point in those five hours, have at least one experience of ethereal love – even if only for a short duration. In fact, I think we can go further than this: if one spends five hours diligently practising loving-kindness meditation, giving the practice one's full attention, giving it one's all, then one is highly likely to have an experience of ethereal love; it would be unusual for one *not* to have an experience of ethereal love during that time.

Recall that under enlightened constructivism, as it currently stands, *X* is a moral reason for *A* to *Y* iff *X* is logically and instrumentally entailed from within the evaluative standpoint of *A*'s peak self at a particular moment in time, the evaluative standpoint being composed of the *full set* of hypothetical attitudes of valuing held by *A*'s peak self – in either an occurrent or dispositional form – together with any non-normative facts that have a bearing on those attitudinal valuations.

As discussed in the preceding sections, one's peak self is a hypothetical version of oneself who is stipulated to have all of the same information, autobiographical memories, etc., as one's current, actual self but who inhabits the same lucid, expansive, joyful, loving, etc., state of mind that one did during one's highest peak experience. One's 'highest' peak experience – at least for the purposes of this chapter – is that which, on balance, is the *most* joyous and which involves the *greatest* feeling of clarity, insight, and higher-order stability (higher-order stability with respect to any changes in one's evaluations that occur during a peak experience, whether these be changes in the direction of greater coherence or transformative changes – or both).

The modification to the enlightened constructivist method that I am proposing runs as follows. *X* is a moral reason for *A* to *Y* iff *X* is logically and instrumentally entailed from within the evaluative standpoint of *A*'s *most loving self* at a particular moment in time, the evaluative standpoint being composed of the *full set* of hypothetical attitudes of valuing held by *A*'s most loving self – in either an occurrent or dispositional form – together with any non-normative facts that have a bearing on those attitudinal valuations.

What is one's most loving self? It is a hypothetical version of oneself who is stipulated to have all of the same information, autobiographical memories, etc., as one's current, actual self but who inhabits the same transcendentally loving, kind, and compassionate state of mind that one did during one's highest experience of ethereal love (or that one *would do* if one were to follow the methodology of loving-kindness meditation). One's 'highest' experience of ethereal love – at least for the purposes of this chapter – is that which, on balance, is the *most* loving, kind, and compassionate and which involves the *greatest* feeling of clarity, insight, and higher-order stability (higher-order stability with respect to any changes in one's evaluations that occur during an experience of ethereal love, whether these be changes in the direction of greater coherence or transformative changes – or both).[35] I will call this view 'meditative constructivism'.

I have already sketched out one theory-independent reason for favouring meditative constructivism over enlightened constructivism, which is that experiences of ethereal love tend to be more morally insightful than conventional peak experiences, which makes them more relevant to the meta-ethical constructivist method for generating truths about one's moral reasons. What makes this point a bit confusing is that constructivism doubles up as both a meta-ethical theory and a meta-normative theory: it offers an account of what makes statements about one's moral reasons true and also an account of what makes statements about one's normative practical reasons true more generally.

As discussed in Chapter 1, I am interested in defending constructivism as a theory of ethical truth rather than as a theory of normative truth. The fact that experiences of ethereal love tend to be more morally insightful, in general, than

---

[35] Recall my example of Heinrich and Matthias.

conventional peak experiences is relevant when it comes to the *meta-ethical* constructivist method but not when it comes to the meta-normative constructivist method.

A second independent reason for favouring meditative constructivism over enlightened constructivism would be if ethical truth were more accessible, more knowable, under the former than under the latter. This rests on the assumption that *other things being equal*, one meta-ethical theory is preferable to another if, under that theory, truths about what moral reasons an agent possesses are significantly more accessible to that agent (as covered in Chapter 1, I am focusing solely on moral reasons in this thesis; I do not discuss other facets of moral truth, such as moral obligations). Whatever our moral reasons, in fact, are, it is preferable to know *what* they are, as opposed to having that moral knowledge remain shrouded in mystery.

The advantage of meditative constructivism is that it involves a systematic methodology – i.e. loving-kindness meditation – for eliciting the ethically salient state: the state of ethereal love. If one proceeds far enough with this methodology, one is highly likely to have an experience of ethereal love. The general idea here is that this feature of the theory would allow an agent to imaginatively occupy the position of their *most loving self* with greater accuracy and frequency.

So, if one is ever in doubt as to what evaluations one's most loving self *would* hold, one can sit down, close one's eyes, and practise loving-kindness meditation; if one practises diligently enough and for long enough, one of two things will happen: (a) one will elicit a *new* highest state of ethereal love (i.e. a more intense experience of ethereal love than one has ever undergone hitherto); (b) one will elicit a state of ethereal love that is *not* one's highest but which is sufficiently similar to one's highest that it allows one to approximate the evaluative perspective of one's most loving self.

In the case of (a), for as long as one occupies that new highest state of ethereal love, one is *identical* to one's most loving self: there is no experiential distance between one's actual self and one's most loving self. When this happens, one does not have to speculate as to what evaluations one's most loving self would hold because one will simply *be* that self (for a short duration). One can use the methodology of loving-kindness meditation to create a temporary *experiential bridge* between one's actual self and one's most loving self, the hypothetical being whose

attitudes go towards fixing the truths about what moral reasons one possesses. And one can create this experiential bridge over and over again, as many times as one is willing to practise loving-kindness meditation.[36]

In the case of (b), one is not quite experientially *identical* to one's most loving self, but one is close enough that one can imaginatively occupy the position of this hypothetical being with greater accuracy ('greater' relative to mundane periods of waking consciousness). My tacit assumption here is that diverse experiences of the state of ethereal love will be very similar both in phenomenological character and in the changes that they precipitate in one's subjective motivational set. I think this is a reasonable assumption given that the key ingredients of the state of ethereal love do not vary much, if at all, between each instantiation of it. The core of an experience of ethereal love almost always consists of a beatific harmony, a synaesthetic triptych, of the three moral emotions of love, kindness, and compassion. And these moral emotions, when raised to a fever pitch of intensity, will generally strengthen one's pro-social, cooperative, altruistic, and other-regarding evaluations.

There are some exceptions to this – possible, if not actual. An agent might experience extraordinary love and kindness while practising loving-kindness meditation but *not* experience extraordinary compassion – so only two out of the three emotions that comprise the triptych are present. And this might, in turn, have a dampening effect on the evaluative changes that are wrought by the experience. But if such an experience satisfies all of the other conditions stated in my definition of peak experiences (see §4.7), then we might concede that it still counts as an experience of ethereal love (recall that ethereal love is just a particular kind of peak experience).

The key problem with what I have said so far is that gaining a better understanding of what evaluations one's most loving self would hold does not *necessarily* translate into gaining a better understanding of what one's moral reasons are. This is because

---

[36] To be clear, I am not suggesting that loving-kindness meditation is the *only* means by which to kindle an experience of ethereal love; one could, presumably, stumble into such experiences throughout one's life in the same way that one stumbles into conventional peak experiences (e.g. while walking in the countryside on a fine Summer day). But loving-kindness meditation is a reliable and systematic means for eliciting experiences of ethereal love, and that is why I am focusing on it to the exclusion of non-systematic means.

undergoing an experience of ethereal love will not necessarily give one a better understanding of what the *relevant non-normative facts* are. I discussed at length in Chapter 3, §3.8, the crucial and often decisive role that non-normative facts play in determining the nature of one's moral reasons. To recycle an example that I used in that section, if, unbeknownst to oneself, one's boiler is leaking poisonous carbon monoxide gas, putting oneself and one's family at risk, then it seems plausible to say that one has most moral reason to immediately evacuate the house or apartment where one is living before inhalation of the gas results in death by suffocation. Just for some context, carbon monoxide gas is both odourless and colourless, which makes it especially dangerous in such situations (Britannica Academic 2023: 'carbon monoxide poisoning').

So, let us say that one is sitting there in one's house or apartment, practising loving-kindness meditation, all the while blissfully unaware that one's boiler is leaking toxic carbon monoxide gas. Even if one succeeds in eliciting the most intense experience of ethereal love that one has ever undergone in one's life, and even if this experience grants one a heightened understanding of what evaluations one's *most loving self* would hold, it still leaves one ignorant as to what one's overriding moral reason is in this case, which is to get the hell out of the house before the carbon monoxide gas proves lethal. The methodology of loving-kindness meditation can help one to illuminate the evaluations that go towards fixing ethical truth, but it cannot play this same luminary role when it comes to the relevant non-normative facts that go towards fixing ethical truth.

To recycle another example from Chapter 3, loving-kindness meditation cannot inform one whether one's close relative is suffering from an undiagnosed brain tumour or not (as discussed in §3.8, a neurological condition of this kind would most likely have a bearing on one's evaluations concerning that individual). One important caveat here is that during loving-kindness meditation one can sometimes form new conceptual connections or discern new patterns between certain existing bits of information that one has about the world. So, for instance, if one has noticed a gradual deterioration in the moral behaviour of a close relative over the course of several months, a deterioration that is completely out of keeping with that person's character, then one might, in a sudden flash of insight while practising loving-kindness meditation that is directed towards that particular individual, realise that

there may be a medical cause underlying the behavioural deterioration, such as an undiagnosed brain tumour.

In sum, I do not think it is plausible to say that ethical truth is significantly more accessible under meditative constructivism than under enlightened constructivism; the most we can say is that the evaluations that go towards fixing ethical truth are more accessible. This result is better than nothing – it is certainly better than those attitudes becoming *less* accessible – but it is not quite what was hoped for, which was a significant increase in the accessibility of truths about what moral reasons one has.

So, out of the two independent motives that I have sketched out for exploring a meditation-based modification to enlightened constructivism, only one turns out to be plausible, which is that experiences of ethereal love tend to have a more obvious ethical dimension to them, in general, than conventional peak experiences. Let us now turn to the question of whether meditative constructivism can muster a more compelling response to the perverse peak experiences objection than enlightened constructivism.

4.9 GENGHIS KHAN, TAKE TWO

Our long, meandering detour into the land of meditation has come full circle: we are now back where we started, with an objection that I first raised in §4.4. But this long detour has not been without purpose: we have 'levelled up' from enlightened constructivism to meditative constructivism – just as, in the video game *Pokémon*, one might 'level up' a Pokémon creature so that it becomes more powerful, more capable of holding its own against rival Pokémon creatures (i.e. either rival meta-ethical theories or strong objections to this particular theory, under this analogy). So, let us see what our levelled-up version of constructivism, meditative constructivism, has to say in response to the perverse peak experiences objection. Let us deploy our Pokémon and see whether it can win this fight.

First, let me offer a quick recap of the perverse peak experiences objection. Recall that peak experiences – which include experiences of ethereal love, as elicited through loving-kindness meditation – might prompt two kinds of change in an agent's subjective motivational set: change in the direction of greater coherence and transformative change. As covered in §4.2, the former kind of change involves

identifying and, if necessary, remedying failures of coherence in one's set of values (e.g. an agent might see for the first time how certain values are in tension with one another, might grasp the full implications of certain values, might become suddenly cognisant of values that had hitherto been implicit or buried in dark recesses of her mind, etc.). The latter kind of change, by contrast, involves substantive changes in one's values that are higher-order stable and which go beyond merely remedying failures of coherence. Such substantive changes might include the formation of a dozen new strongly-held evaluations in one's motivational set, say.

As discussed in §4.4, change in the direction of greater coherence is the sort of thing that can be measured – at least *in principle* – in an 'objective', stance-independent way. That is, we can measure it with reference to standards of coherence that are *independent* of the evaluations of any particular agent. For example, if two particular evaluative judgments are in logical or practical contradiction with one another, then this fact – the bare fact of their being in contradiction – does not *itself* obtain in virtue of the contingent attitudes of any particular agent. But the same cannot be said of transformative change. There is no stance-independent way, under a constructivist view of things, of judging whether a radical, substantive change in an agent's values during a peak experience represents a genuine improvement in that agent's values.

The original version of the perverse peak experiences objection is aimed at *conventional* peak experiences rather than experiences of ethereal love; I will adapt it to the case of ethereal love a bit further on. But to recap the original objection, it is that we can think of conventional peak experiences that substantively change an agent's values in a way that seems, from his own, first-person point of view, to have significantly *improved* his values but which, from the point of view of other agents, seems to have achieved the exact opposite. That is, a peak experience could change an agent's moral outlook in a way that is higher-order stable from the perspective of that agent but which seems utterly perverse and deranged from the perspective of other agents. To illustrate this point, I quoted an extraordinary pronouncement from the great Mongol Emperor Genghis Khan: 'The greatest joy a man can know is to conquer his enemies and drive them before him. To ride their horses and take away their possessions. To see the faces of those who were dear to

them bedewed with tears, and to clasp their wives and daughters in his arms' (Gat 2006: 427).

In §4.4 I sketched out a possible peak experience, inspired by this quote, that Genghis Khan might have had – or could very easily *have* had. I will briefly restate it here. We can imagine Genghis Khan standing on a hilltop, gazing out over the smouldering remains of a city he has just plundered, and being flooded by a feeling of preternatural joy and clarity. It is not hard to imagine such an experience seeming, *from his own, first-person point of view*, to communicate profound moral insights. As he becomes overwhelmed by joy, the incomparable joy of conquest, the great Khan might think to himself: *Yes, this is exactly what I was born to do. I am fulfilling my destiny. I am a father to the Mongol people, and I must forge a new realm for them. To do this, I need to topple the existing order, to raze cities to the ground and enslave their citizens, to bend proud nations to my will. There can be no room for pity or weakness. If I am to walk the path that fate has cut out for me, if I am to build a new world for my people, I must first destroy the existing one. There will be terrible suffering, but it will all be for the good in the end.*

Intuitively, we feel that any agent who judges himself to have moral reasons to engage in ruthless, genocidal military conquest is simply mistaken, and this is true even if the agent in question has arrived at these judgments via an experience of transcendent joy, clarity, insight, etc. In other words, we feel that Genghis Khan's set of values has *not* been improved by this peak experience, and this is true no matter how intense Genghis Khan's subjective feeling of heightened clarity, insight, and higher-order stability was while he underwent the experience. The problem is that *enlightened* constructivism – as opposed to *meditative constructivism*, which I will get to shortly – is unable to accommodate this intuition. The enlightened constructivist cannot rule out the possibility of perverse peak experiences, such as Genghis Khan's. If a peak experience causes seemingly perverse changes in an agent's values, there is nowhere to stand *outside* of any individual agent's evaluative standpoint from which to say that that particular agent's values have worsened as a result of those changes – they just seem like changes for the worse from the point of view of one's *own* evaluative standpoint (or someone else's, or some combination of particular evaluative standpoints).

I argued in §4.4 that the best that the enlightened constructivist can do with respect to this objection is to try to show that perverse peak experiences are, in reality, vanishingly rare and that peak experiences almost always change an agent's values in a way that seems, *both from that agent's own point of view and from the point of view of other agents*, like a significant, higher-order-stable improvement (though, once again, there is no stance-independent way of establishing that it does, in fact, constitute an improvement). Unfortunately, enlightened constructivism does not have the resources to do even this much. I discussed towards the end of §4.4 the main reason for this, which is that the emotion of very intense joy, which chiefly characterises conventional peak experiences, can be had in so many contexts that seem pathological.

The hope was that meditative constructivism, with its emphasis on experiences of ethereal love rather than conventional peak experiences, would be able to narrow the range of possible perverse peak experiences that one could have under the view, thereby rendering their occurrence vastly less likely in the actual world. The argumentative strategy here is not to try and rule out the bare possibility of an agent having moral insights during a peak experience that seem utterly perverse and deranged from the perspective of other agents but to try and show that this possibility is not as alarming or revisionist as it might, at first blush, seem (because it is very, very rarely actualised). In other words, the strategy here is to try and show that the revisionist element of the theory is tolerably rather than intolerably revisionist.

So, would an agent be significantly less likely to have a perverse experience of ethereal love, as elicited through loving-kindness meditation, than a conventional peak experience? I think the answer to this question is 'yes', and the reason for it comes down to the difference in the emotions involved in these experiences. As discussed in §4.7, the core of an experience of ethereal love consists of a harmonious triptych of three moral emotions that have been amplified to an extraordinary degree of intensity: love, kindness, and compassion. What makes these emotions *moral* is that they are 'linked to the interests or welfare either of society as a whole or at least of persons other than the judge or agent' (Haidt 2003: 853). In other words, they have a strong other-regarding dimension to them.

It is relatively easy to think of cases where the emotion of very intense joy that is characteristic of regular, non-meditative peak experiences seems perverse, pathological, sadistic, etc. It is much harder to do this when it comes to the emotions of extraordinary love, kindness, and compassion that are characteristic of experiences of ethereal love. Let us focus on the emotion of kindness for a moment. Kindness is about being 'benevolent, friendly, or warm-hearted' in 'nature or disposition'; it is about being 'ready to assist, or show consideration for, others'; it involves being 'sympathetic, obliging, considerate'; a kind action is one which expresses 'generous, caring, or sympathetic thoughts or feelings' (OED Online 2023: 'kind, adj. and adv.'). With this in mind, it is genuinely hard to think of cases where the emotion of very intense kindness seems perverse, pathological, sadistic, etc. (or, more to the point, where it results in changes in one's evaluations that seem this way). Any notion of 'sadistic kindness', for instance, seems like an oxymoron.

The same is true for both compassion and love. These three emotions, kindness, compassion, and love, are all about having a benevolent and warm-hearted orientation towards other living beings. These emotions entail that we be concerned with the well-being and flourishing of others (though not necessarily *human* others: one might be concerned with the well-being and flourishing of non-human animals, say). With this in mind, it seems highly probable that if we massively amplify these three emotions during loving-kindness meditation, then any transformative changes in one's evaluations that occur as a result of that amplification will be benevolent in character. It would be an astonishing aberration for an agent to be flooded with feelings of transcendent love, kindness, and compassion and then to have this experience change their evaluations in such a way as to make them more selfish, other-denying, or cruel, say.

To bring this back to Genghis Khan, we can easily imagine Genghis Khan – or someone with a mind very similar to Genghis Khan – being flooded with transcendent joy, and this changing his evaluations in a way that seems perverse. But if Genghis Khan were, hypothetically, to practise loving-kindness meditation and elicit an experience of ethereal love, during which time he was overwhelmed by feelings of transcendent love, kindness, and compassion, then it is far harder to imagine this experience changing his evaluations in a way that seems perverse. Again, these emotions are all about being benevolently and warm-heartedly

disposed towards other living beings, which makes it far less likely that an agent would glean perverse moral insights from any meditative experience centred around them.

So, meditative constructivism would appear to have achieved the desired result, which is to vastly reduce the likelihood of an agent having a perverse peak experience. But there is one huge, glaring problem with all that has been said so far, and that is that love, kindness, and compassion are not the *only* moral emotions. Recall that what makes an emotion *moral* is that it is 'linked to the interests or welfare either of society as a whole or at least of persons other than the judge or agent' (Haidt 2003: 853). In other words, moral emotions have a strong other-regarding component to them. But just because a moral emotion has to be other-regarding does not mean that it has to be other-regarding in a *benevolent* way. For instance, Haidt (2003: 854) describes anger and guilt as two 'prototypical moral emotions'. He goes on to say that '[e]motions that motivate helping behavior are easy to label as moral emotions, but emotions that lead to ostracism, shaming, and murderous vengeance are no less a part of our moral nature' (ibid.: 855). An emotion like anger, say, almost always concerns others: one might feel anger at government ministers for introducing some new policy that one deems to be foolish and misguided, anger at a terrorist in the news who has ruined the lives of many, anger at a friend for having let one down, anger at poachers for illegally hunting endangered species, anger at corrupt officials who have abused their positions of power, etc. There are certain cases in which anger can be purely self-regarding, such as when one is angry at oneself for having flubbed an important task, but such cases are relatively rare.

This raises a very uncomfortable question for the meditative constructivist: why are we focusing exclusively on states of ethereal love that are elicited through loving-kindness meditation in order to construct truths about our moral reasons when there are moral emotions *other* than love, kindness, and compassion that are highly relevant to any method for generating ethical truth? Why couldn't there be a form of meditation that mirrored loving-kindness meditation in all respects except that it focused on eliciting and cultivating states of transcendent anger, guilt, or gratitude instead of states of ethereal love? Recall from §4.8 that one of the key independent reasons for drawing on experiences of ethereal love in the constructivist method was

that states of ethereal love are, in general, more *morally insightful* than conventional peak experiences. But this raises the question of whether they are *also* more morally insightful than every other variety of (hypothetical) meditative experience involving the amplification of moral emotions; they would *have* to be in order for us to be at all justified in focusing on experiences of ethereal love to the exclusion of all other ethically salient meditative experiences.

Let us make this more concrete by considering a hypothetical alternative to loving-kindness meditation. We can imagine Genghis Khan engaging in *ethereal-rage meditation*. Imagine that Genghis Khan finds a quiet corner of his grand emperor's tent, closes his eyes, and begins by directing his attention towards feelings of anger for a particular person. In this case, it is the governor of a fortified enemy city who is stubbornly resisting the attempts of the Mongol army to capture the city. To help focus his attention, Genghis Khan silently repeats the following mantra to himself (directed at the governor). *May you be wretched; may you feel pain; may your life be filled with sorrow*. Genghis Khan slowly kindles these feelings of anger until they expand to encompass not just the governor but everyone living within the city's walls. He is wroth at all of them for daring to resist his imperium. Let us say that Genghis Khan eventually succeeds in eliciting a state of *ethereal rage*, wherein an 'unconditional feeling of … [anger] pervades the whole mind as a way of being, with no other consideration, or discursive thoughts' (Lutz et al. 2008: 1). This rage is pristine and untrammelled, as perfect as a cut gemstone; it is a rage that burns with the purity and intensity of a white-hot flame, the kind of flame that blacksmiths use in their forges for heating metal. When in this state, it is as though Genghis Khan has transcended his humanity. He is no longer a man but the personification of an emotion: the emotion of anger. He is like the *god* of anger, the kind of god that warriors might pray to.

We can imagine this state of ethereal rage sharing most of the key features of a state of ethereal love, with the major exception being that it is characterised by the moral emotion of anger rather than the moral emotions of love, kindness, and compassion. We can imagine an experience of ethereal rage seeming, from Genghis Khan's perspective, like one of greatly heightened clarity and moral insight relative to periods of mundane waking consciousness. We can imagine it prompting transformative changes in his evaluations that are higher-order stable from his

perspective but which seem utterly perverse and deranged from the perspective of other agents. In short, we can see that meditative constructivism has not solved the problem of perverse peak experiences; all it has done is transfigured the problem into one of perverse *meditative* experiences.

The difficulty faced by meditative constructivism is that there is no principled, non-*ad hoc* way of excluding perverse meditative experiences from the constructivist method. Any meditative experiences that involve the amplification of moral emotions – whether these be emotions of anger, guilt, compassion, disgust, shame, contempt, gratitude, etc. – are very likely going to be ethically salient in the sense of altering those evaluations that relate to other agents or to one's interactions with other agents (i.e. those evaluations that are other-regarding in character). And assuming that these alternative meditative experiences involve – or, at least, sometimes involve – a sense of greatly heightened clarity and moral insight on the part of the individual experiencing them, then they are, by the constructivist's logic, going to be relevant to any hypothetical, meta-ethical constructivist method for generating truths about one's moral reasons.

Meditative constructivism focuses on a subset of the full set of hypothetical meditative experiences that involve moral emotions: it focuses on experiences of ethereal love that are elicited through loving-kindness meditation. Obviously, by narrowing the focus of the theory in this way, we make it more likely that we get a desirable result ('desirable' from the point of view of making constructivism a more attractive meta-ethical theory). This is because, as argued earlier in this section, experiences of ethereal love almost always change an agent's evaluations in a way that is *benevolent*, and this, in turn, will generally result in a strengthening of the moral reasons that a given agent possesses (depending on what the relevant non-normative facts are in that particular agent's case). But the problem is that there are hypothetical meditative experiences, such as experiences of *ethereal rage*, that could easily appear from the point of view of the agent undergoing them to be just as morally insightful, if not more so, than experiences of ethereal love and yet which change that particular agent's evaluations in a way that seems *malevolent* rather than benevolent. If such experiences are ethically salient, which they are, then we cannot exclude them from the constructivist method simply because they yield an undesirable result.

<center>***</center>

In closing, I think we can say that while meditative constructivism represents an interesting position in logical space, the theory faces significant unanswered challenges relating to perverse meditative experiences – and the same goes for enlightened constructivism and perverse peak experiences. To be clear, I am not claiming that there is *no* possible version of meditative constructivism that could be made to work; there may be some variant of the theory floating out there in logical space that *is* able to surmount the various challenges that I have raised in this chapter. But an analysis of any such variant is beyond the scope of this thesis.

In the introduction to this thesis I compared myself to a cartographer mapping out an underground cave system. In this analogy, the cave system stands for the network of meta-ethical views that fall under the heading of 'constructivism'. The main subterranean cave that I have been poking around in throughout this thesis is that relating to *Humean* constructivism; my discussion of the theory formed the principal subject matter of chapters 1 through 3. *Enlightened* constructivism, which I evaluated in §4.3 and §4.4 of this chapter, is like a smaller cave branching off from the main cave that is Humean constructivism. And *meditative* constructivism, which I introduced in §4.8, is, in turn, like an even smaller cave branching off from the sub-cave that is enlightened constructivism: it is a cave within a cave within a cave.

We have seen many interesting things in the course of this subterranean exploration. We have seen stalactites hanging from the ceiling and stalagmites rising up from the floor. We have seen gemstone ore lodged in the rocky walls. We have seen underground lakes, abandoned mineshafts, and strange subterranean flora. And in the last of these caves, the cave that is meditative constructivism, we even caught sight of a troglodyte lurking in the shadows (which represents the perverse meditative experiences objection in this analogy). The troglodyte was unwilling to yield its territory, unwilling to let us pass further into this underground cave system, and rather than try to fight it, we chose to back slowly out of the cave: to backtrack all the way back to the main cave that is Humean constructivism.

And that is where I will end this thesis. In Chapter 1 I prodded and probed Humean constructivism from every angle, comparing it to rival forms of constructivism; in Chapter 2 I evaluated its strengths; in Chapter 3 I evaluated its weaknesses; in Chapter 4, this chapter, I explored two interesting off-shoots of

Humean constructivism: enlightened constructivism and meditative constructivism. The map that I have made of this underground cave system is by no means complete, but it is a start. Hopefully, it may prove of some use to those who are interested in defending, developing, understanding, or criticising a constructivist view of ethical truth.

BIBLIOGRAPHY

Aaltola, E. (2015) 'Wilderness Experiences as Ethics: From Elevation to Attentiveness', *Ethics, policy & environment*, 18(3), pp. 283–300. doi:10.1080/21550085.2015.1111614.

'Achilles' (2023) *Britannica Academic*. Retrieved 30 March 2023, from https://academic-eb-com.nottingham.idm.oclc.org/levels/collegiate/article/Achilles/3532.

Albahari, M. (2009) 'Witness-Consciousness: Its Definition, Appearance and Reality', *Journal of Consciousness Studies*, 16(1), pp. 62–84.

— 2010. 'Nirvana and Ownerless Consciousness', in Siderits, M., Thompson, E., and Zahavi, D. (eds) *Self, No Self?: Perspectives from Analytical, Phenomenological, and Indian Traditions*. Oxford: Oxford University Press, pp. 79–113.

Alcock, J. (2013) *Animal behavior : an evolutionary approach*. 8th ed. Sunderland, Mass.: Sinauer Associates.

Alston, W.P. (1996) *A realist conception of truth*. Ithaca: Cornell University Press.

Ananth, M. (2005) 'Psychological altruism vs. biological altruism: Narrowing the gap with the Baldwin effect', *Acta biotheoretica*, 53(3), pp. 217–239. doi:10.1007/s10441-005-2530-6.

Bagnoli, C. (2013) 'Introduction', in Bagnoli, C. (ed.) *Constructivism in ethics*. New York: Cambridge University Press, pp. 1–21.

— (2021) 'Constructivism in Metaethics', in Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition). Available at: https://plato.stanford.edu/archives/spr2021/entries/constructivism-metaethics/.

Bankard, J. (2015) 'Training Emotion Cultivates Morality: How Loving-Kindness Meditation Hones Compassion and Increases Prosocial Behavior', *Journal of religion and health*, 54(6), pp. 2324–2343. doi:10.1007/s10943-014-9999-8.

Batson, C.D. *et al.* (2005) 'Empathy and Altruism', in Snyder, C. R. and Lopez, S. J. (eds) *Handbook of Positive Psychology*. New York: Oxford University Press, pp. 485–498.

Behrends, J. (2013) 'Meta-normative Realism, Evolution, and Our Reasons to Survive', *Pacific philosophical quarterly*, 94(4), pp. 486–502. doi:10.1111/papq.12010.

— (2017) 'Normative Source and Extensional Adequacy', *Journal of ethics & social philosophy*, 10(3), pp. 1–26. doi:10.26556/jesp.v10i3.103.

Berkowitz, L. (1972) 'Social Norms, Feelings, and Other Factors Affecting Helping Behavior and Altruism', in Berkowitz, L. (ed.). *Advances in Experimental Social Psychology*. New York: Academic Press, pp. 62–106.

Bethelmy, L.C. and Corraliza, J.A. (2019) 'Transcendence and Sublime Experience in Nature: Awe and Inspiring Energy', *Frontiers in psychology*, 10, pp. 509–509. doi:10.3389/fpsyg.2019.00509.

Björklund, F. *et al.* (2012) 'Recent Work on Motivational Internalism', *Analysis (Oxford)*, 72(1), pp. 124–137. doi:10.1093/analys/anr118.

Björnsson, G. *et al.* (2015) 'Introduction', in Björnsson, G. *et al.* (eds). *Motivational Internalism*. Oxford: Oxford University Press, pp. 1–20.

Blackburn, S. (1993) *Essays in quasi-realism*. New York: Oxford University Press.

— (1998) *Ruling passions : a theory of practical reasoning*. Oxford: Clarendon Press.

— (2010) 'The Majesty of Reason', *Philosophy (London)*, 85(1), pp. 5–27. doi:10.1017/S0031819109990428.

Bloom, H. (2004) *The best poems of the English language : from Chaucer through Robert Frost*. New York: HarperCollins Publishers.

Bloom, P. (2018) *Against empathy : the case for rational compassion*. New York: HarperCollins Publishers.

Bodian, S. (2017) *Beyond mindfulness : the direct approach to lasting peace, happiness, and love*. Oakland, California: New Harbinger Publications.

Brink, D.O. (1986) 'EXTERNALIST MORAL REALISM', *The Southern journal of philosophy*, 24(S1), pp. 23–41. doi:10.1111/j.2041-6962.1986.tb01594.x.

— (1989) *Moral realism and the foundations of ethics*. Cambridge: Cambridge University Press.

— (1997) 'Moral Motivation', *Ethics*, 108(1), pp. 4–32. doi:10.1086/233786.

Buss, D.M. (2011) *Evolutionary psychology : the new science of the mind*. 2nd ed. Boston, Mass.: Allyn & Bacon.

Carbon monoxide poisoning (2023) *Britannica Academic*. Retrieved 21 June 2023, from https://academic-eb-com.nottingham.idm.oclc.org/levels/collegiate/article/carbon-monoxide-poisoning/100872.

Chaucer, G. (1985) *The Canterbury Tales*. Oxford: Oxford University Press.

Chappell, S.G. (2022) *Epiphanies : an ethics of experience*. Oxford: Oxford University Press.

Cholbi, M. (2006) 'Belief Attribution and the Falsification of Motive Internalism', *Philosophical psychology*, 19(5), pp. 607–616. doi:10.1080/09515080600901939.

Chu, L-C. (2010) 'The benefits of meditation vis-à-vis emotional intelligence, perceived stress and negative mental health', *Stress and health*, 26(2), pp. 169–180. doi:10.1002/smi.1289.

Copp, D. (1991) 'Moral Realism: Facts and Norms', *Ethics*, 101(3), pp. 610–624. doi:10.1086/293320.

Cordner, C. (2016) 'Unconditional love?', *Cogent arts & humanities*, 3(1), doi:10.1080/23311983.2016.1207918.

Dancy, J. (2006) 'What Do Reasons Do?', in Horgan, T. and Timmons, M. (eds). *Metaethics after Moore*. Oxford: Oxford University Press, pp. 36–90.

Darwall, S., Gibbard, A., and Railton, P. (1992) 'Toward Fin de siècle Ethics: Some Trends', *The Philosophical review*, 101(1), pp. 115–189. doi:10.2307/2185045.

Darwin, C. (1874) *The descent of man, and selection in relation to sex*. New York: Rand, McNally & Company.

Dawkins, R. (2006) *The selfish gene*. 30th anniversary ed. Oxford: Oxford University Press.

— (1996) *Climbing Mount Improbable*. London: Penguin Books.

Deonna, J. and Teroni, F. (2022) 'Why are emotions epistemically indispensable?', *Inquiry* (Oslo), ahead-of-print, pp. 1–23. doi:10.1080/0020174X.2022.2126147.

Dorsey, D. (2018) 'A Perfectionist Humean Constructivism', *Ethics*, 128(3), pp. 574–602. doi:10.1086/696152.

— (2019) 'Preferences and Prudential Reasons', *Utilitas*, 31(2), pp. 157–178. doi:10.1017/S0953820818000250.

Dostoevsky, F. (1994) *The Karamazov brothers*, translated by Ignat Avsey. Oxford: Oxford University Press.

Enoch, D. (2009) 'Can there be a global, interesting, coherent constructivism about practical reason?', *Philosophical explorations*, 12(3), pp. 319–339. doi:10.1080/13869790903067683.

Finlay, S. (2006) 'The reasons that matter', *Australasian journal of philosophy*, 84(1), pp. 1–20. doi:10.1080/00048400600571661.

Firth, R. (1952) 'Ethical Absolutism and the Ideal Observer', *Philosophy and phenomenological research*, 12(3), pp. 317–345. doi:10.2307/2103988.

Flaubert, G. (1992) *Madame Bovary*, translated by Francis Steegmuller. New York: Vintage Books.

Foot, P. (2002) *Virtues and vices and other essays in moral philosophy*. Oxford: Clarendon.

Forcehimes, A.T. and Semrau, L. (2018) 'Are There Distinctively Moral Reasons?', *Ethical theory and moral practice*, 21(3), pp. 699–717. doi:10.1007/s10677-018-9919-1.

Fredrickson, B.L. *et al.* (2008) 'Open Hearts Build Lives', *Journal of personality and social psychology*, 95(5), pp. 1045–1062. doi:10.1037/a0013262.

'Fyodor Dostoyevsky' (2023) *Britannica Academic*. Retrieved 30 March 2023, from https://academic-eb-com.nottingham.idm.oclc.org/levels/collegiate/article/Fyodor-Dostoyevsky/108410.

Garfield, J. (2022) *Losing ourselves : learning to live without a self*. Princeton: Princeton University Press.

Gaskin, R. (2018) *Tragedy and redress in Western literature*. London: Routledge.

Gat, A. (2006) *War in human civilization*. New York: Oxford University Press.

'Genghis Khan' (2023) *Britannica Academic*. Retrieved 26 April 2023, from https://academic-eb-com.nottingham.idm.oclc.org/levels/collegiate/article/Genghis-Khan/109425.

Gibbard, A. (1999) 'Morality as Consistency in Living: Korsgaard's Kantian Lectures', *Ethics*, 110(1), pp. 140–164. doi:10.1086/233207.

Gill, M.B. (2014) *Humean moral pluralism*. Oxford: Oxford University Press.

Glock, H.-J. (2009) 'Concepts: Where Subjectivism Goes Wrong', *Philosophy (London)*, 84(1), pp. 5–29. doi:10.1017/S0031819109000011.

Godfrin, K.A. and van Heeringen, C. (2010) 'The effects of mindfulness-based cognitive therapy on recurrence of depressive episodes, mental health and quality of life: A randomized controlled study', *Behaviour research and therapy*, 48(8), pp. 738–746. doi:10.1016/j.brat.2010.04.006.

Goldman, A.H. (2006) 'Desire Based Reasons and Reasons for Desires', *The Southern journal of philosophy*, 44(3), pp. 469–488. doi:10.1111/j.2041-6962.2006.tb00014.x.

Goldstein, J. (2013) *Mindfulness : a practical guide to awakening*. Boulder, Colorado: Sounds True.

Gouldner, A.W. (1960) 'The Norm of Reciprocity: A Preliminary Statement', *American sociological review*, 25(2), pp. 161–178. doi:10.2307/2092623.

Goyal, M. *et al.* (2014) 'Meditation Programs for Psychological Stress and Well-being: A Systematic Review and Meta-analysis', *JAMA internal medicine*, 174(3), pp. 357–368. doi:10.1001/jamainternmed.2013.13018.

Greenough, P. and Lynch, M.P. (2006) *Truth and realism*. Oxford: Clarendon.

Haidt, J. (2003) 'The Moral Emotions', in Davidson, R. J., Scherer, K. R., and Goldsmith, H. H. (eds). *Handbook of affective sciences*. Oxford: Oxford University Press, pp. 852–870.

Harari, Y. N. (2011) *Sapiens : a brief history of humankind*. London: Penguin Books.

— (2017) *Homo deus : a brief history of tomorrow*. London: Vintage.

Hare, R.M. (1952) *The language of morals*. Oxford: Clarendon Press.

Harris, S. (2006) *The end of faith : religion, terror, and the future of reason*. London: Free Press.

— (2012) *The moral landscape : how science can determine human values*. London: Black Swan.

— (2014) *Waking up : searching for spirituality without religion*. London: Transworld Publishers.

Harth, M. (2020) 'Faultless Disagreement, Realism and Moral Objectivity', *Erkenntnis*, 85(5), pp. 1085–1108. doi:10.1007/s10670-018-0066-9.

Hasenkamp, W. *et al.* (2012) 'Mind wandering and attention during focused meditation: A fine-grained temporal analysis of fluctuating cognitive states', *NeuroImage (Orlando, Fla.)*, 59(1), pp. 750–760. doi:10.1016/j.neuroimage.2011.07.008.

Hauser, M.D. (2006). *Moral minds : how nature designed our universal sense of right and wrong*. New York: Harper Collins.

Hölzel, B.K. *et al.* (2011) 'Mindfulness practice leads to increases in regional brain gray matter density', *Psychiatry research*, 191(1), pp. 36–43. doi:10.1016/j.pscychresns.2010.08.006.

Hopster, J. (2017) 'Two Accounts of Moral Objectivity: from Attitude-Independence to Standpoint-Invariance', *Ethical theory and moral practice*, 20(4), pp. 763–780. doi:10.1007/s10677-017-9796-z.

Ibsen, H. (1981) *Four major plays*, translated by James McFarlane and Jens Arup. Oxford: Oxford University Press.

Jackendoff, R. (1995) *Languages of the mind : essays on mental representation*. Cambridge: MIT Press.

James, W. (1890) *The Principles of Psychology* (Vol. 1). New York: Henry Holt and Company.

Johnson, B. (2002) 'On the spiritual benefits of wilderness', *International journal of wilderness*, 8(3), pp.28-32.

Joyce, J. (1992) *Dubliners*. London: Penguin Books.

Kawall, J. (2006) 'On the Moral Epistemology of Ideal Observer Theories', *Ethical theory and moral practice*, 9(3), pp. 359–374. doi:10.1007/s10677-006-9016-8.

Kearney, D.J. *et al.* (2014) 'Loving-Kindness Meditation and the Broaden-and-Build Theory of Positive Emotions Among Veterans Disorder', *Medical care*, 52(12), pp. S32–S38. doi:10.1097/MLR.0000000000000221.

Kim, B. *et al.* (2010) 'Effectiveness of a mindfulness-based cognitive therapy program as an adjunct to pharmacotherapy in patients with panic disorder', *Journal of anxiety disorders*, 24(6), pp. 590–595. doi:10.1016/j.janxdis.2010.03.019.

'kind, adj. and adv.' (2023) *OED Online*, Oxford University Press. Retrieved 2 June 2023, from www.oed.com/view/Entry/103445.

Kugel, J.L. (2008) *How to read the Bible : a guide to Scripture, then and now*. New York: Free Press.

Kurland, J.A. and Gaulin, S.J.C. (2005) 'Cooperation and Conflict among Kin', in Buss D. (ed.) *The Handbook of Evolutionary Psychology*. Hoboken, NJ: John Wiley & Sons, pp. 447–482.

Kurzban, R., Burton-Chellew, M.N. and West, S.A. (2015) 'The Evolution of Altruism in Humans', *Annual review of psychology*, 66(1), pp. 575–599. doi:10.1146/annurev-psych-010814-015355.

Korsgaard, C.M. (1996) *The sources of normativity*. Cambridge: Cambridge University Press.

— (2008) *The constitution of agency : essays on practical reason and moral psychology*. Oxford: Oxford University Press.

Krebs, D.L. (2008) 'Morality: An Evolutionary Account', *Perspectives on psychological science*, 3(3), pp. 149–172. doi:10.1111/j.1745-6924.2008.00072.x.

Kriegel, U. (2012) 'Moral Motivation, Moral Phenomenology, And The Alief/Belief Distinction', *Australasian journal of philosophy*, 90(3), pp. 469–486. doi:10.1080/00048402.2011.607464.

Krygier, J.R. *et al.* (2013) 'Mindfulness meditation, well-being, and heart rate variability: A preliminary investigation into the impact of intensive Vipassana meditation', *International journal of psychophysiology*, 89(3), pp. 305–313. doi:10.1016/j.ijpsycho.2013.06.017.

Lampedusa, Giuseppe Tomasi di (2007) *The leopard*, translated by A. Colquhoun. London: Vintage.

Leiberg, S., Klimecki, O. and Singer, T. (2011) 'Short-term compassion training increases prosocial behavior in a newly developed prosocial game', *PloS one*, 6(3), pp. e17798–e17798. doi:10.1371/journal.pone.0017798.

Lenman, J. (2000) 'Consequentialism and Cluelessness', *Philosophy & public affairs*, 29(4), pp. 342–370. doi:10.1111/j.1088-4963.2000.00342.x.

Lenman, J. and Shemmer, Y. (2012) 'Introduction', in Lenman, J. and Shemmer, Y. (eds) *Constructivism in Practical Philosophy*. Oxford: Oxford University Press, pp. 2–17.

'Lion' (2023) *Britannica Academic*. Retrieved 23 May 2023, from https://academic-eb-com.nottingham.idm.oclc.org/levels/collegiate/article/lion/48425.

Lutz, A. *et al.* (2008) 'Regulation of the neural circuitry of emotion by compassion meditation: Effects of meditative expertise', *PloS one*, 3(3), pp. e1897–e1897. doi:10.1371/journal.pone.0001897.

Mackie, J.L. (1990) *Ethics : inventing right and wrong*. London: Penguin Books.

Margolis, E. and Laurence, S. (2021) 'Concepts', in Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition). Available at: https://plato.stanford.edu/archives/spr2021/entries/concepts/.

Martin, G.R.R (2011) *A storm of swords* : *steel and snow*. London: Voyager.

McDonald, M.G., Wearing, S. and Ponting, J. (2009) 'The Nature of Peak Experience in Wilderness', *The Humanistic psychologist*, 37(4), pp. 370–385. doi:10.1080/08873260701828912.

Miller, G.F. (2007) 'Sexual Selection for Moral Virtues', *The Quarterly Review of Biology*, 82(2), pp. 97–125. doi:10.1086/517857.

Nagel, T. (1986) *The view from nowhere*. Oxford: Oxford University Press.

Ñāṇamoli, B. and Bodhi, B. (trans.) (1995) *The middle length discourses of the Buddha*. Somerville, MA: Wisdom Publications.

Olson, E. (1998) 'There is no problem of the self', *Journal of Consciousness Studies*, 5(5–6), pp. 645–657.

Olson, J. (2014) *Moral error theory : history, critique, defence*. First edition. Oxford: Oxford University Press.

Owen, G. E. L. (1960) 'Logic and Metaphysics in Some Earlier Works of Aristotle', in I. Düring and G. E. L. Owen (eds) *Aristotle and Plato in the Mid-Fourth Century* (Studia graeca et latina gothoburgensia, XI). Gothenburg: Almqvist & Wiksell, pp.163–90.

Pater, W. (2008) *Marius the Epicurean*. London: Valancourt Books.

Peach, B. (1953) 'Logical and Practical Contradictions', *Analysis (Oxford)*, 14(2), pp. 43–50. doi:10.1093/analys/14.2.43.

Pinker, S. (2012) *The better angels of our nature : a history of violence and humanity*. London: Penguin Books.

— (2015) *The sense of style : the thinking person's guide to writing in the 21st century*. London: Penguin Books.

— (2022) *Rationality : what it is, why it seems scarce, why it matters*. London: Penguin Books.

Plato. (1981) *Five dialogues*, translated by G.M. A. Grube. Indianapolis: Hackett Publishing Company.

Powell, B. (2013) *The Iliad*. Oxford: Oxford University Press.

Privette, G. (1983) 'Peak experience, peak performance, and flow: A comparative analysis of positive human experiences', *Journal of personality and social psychology*, 45(6), pp. 1361–1368. doi:10.1037/0022-3514.45.6.1361.

Przyrembel, M. and Singer, T. (2018) 'Experiencing meditation – Evidence for differential effects of three contemplative mental practices in micro-phenomenological interviews', *Consciousness and cognition*, 62, pp. 82–101. doi:10.1016/j.concog.2018.04.004.

Railton, P. (1986) 'Moral Realism', *The Philosophical review*, 95(2), pp. 163–207. doi:10.2307/2185589.

Rawls, J. (1980) 'Kantian Constructivism in Moral Theory', *The Journal of philosophy*, 77(9), pp. 515–572. doi:10.2307/2025790.

— (1999) *A theory of justice : revised edition*. Cambridge: Harvard University Press.

Ricard, M., Lutz, A. and Davidson, R.J. (2014) 'Mind of the Meditator', *Scientific American*, 311(5), pp. 38–45. doi:10.1038/scientificamerican1114-38.

Rodrigues, A.M.M. and Gardner, A. (2022) 'Reproductive value and the evolution of altruism', *Trends in ecology & evolution* (Amsterdam), 37(4), pp. 346–358. doi:10.1016/j.tree.2021.11.007.

Rosati, C.S. (2016) 'Moral Motivation', in Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). Available at: https://plato.stanford.edu/archives/win2016/entries/moral-motivation/.

— (2017) 'Mind-Dependence and Moral Realism', in McPherson T. and Plunkett D. (eds) *The Routledge Handbook of Metaethics*. New York: Routledge, pp. 355–370.

Rose, S., Zell, E. and Strickhouser, J.E. (2020) 'The Effect of Meditation on Health: a Metasynthesis of Randomized Controlled Trials', *Mindfulness*, 11(2), pp. 507–516. doi:10.1007/s12671-019-01277-6.

Roskies, A. (2003) 'Are ethical judgments intrinsically motivational? Lessons from "acquired sociopathy"', *Philosophical psychology*, 16(1), pp. 51–66. doi:10.1080/0951508032000067743.

Russell, B. (2004) *History of Western philosophy*. New York: Routledge Classics.

Sarkissian, H. *et al.* (2011) 'Folk Moral Relativism', *Mind & language*, 26(4), pp. 482–505. doi:10.1111/j.1468-0017.2011.01428.x.

Sartre, J-P. (2020) *Nausea*, translated by R. Baldick. London: Penguin Books.

Sayre-McCord, G. (1986) 'The Many Moral Realisms', *The Southern journal of philosophy*, 24(S1), pp. 1–22. doi:10.1111/j.2041-6962.1986.tb01593.x.

— (2017) 'Moral Realism', in Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition). Available at: https://plato.stanford.edu/archives/fall2017/entries/moral-realism/.

Scanlon, T. (1998) *What we owe to each other*. Cambridge, Mass.: Harvard University Press.

Schmitt, N.C. (1994) 'Ecstasy and Peak-Experience: W. B. Yeats, Marghanita Laski, and Abraham Maslow', *Comparative drama*, 28(2), pp. 167–181. doi:10.1353/cdr.1994.0012.

Schooler, J.W. *et al.* (2011) 'Meta-awareness, perceptual decoupling and the wandering mind', *Trends in cognitive sciences*, 15(7), pp. 319–326. doi:10.1016/j.tics.2011.05.006.

Schroeder, M. (2007) *Slaves of the passions*. Oxford: Oxford University Press.

— (2021) 'The fundamental reason for reasons fundamentalism', *Philosophical studies*, 178(10), pp. 3107–3127. doi:10.1007/s11098-020-01572-2.

Segal, N.L., Hershberger, S.L. and Arad, S. (2003) 'Meeting One's Twin: Perceived Social Closeness and Familiarity', *Evolutionary psychology*, 1(1), p. 147470490300100. doi:10.1177/147470490300100105.

Shafer-Landau, R. (2003) *Moral realism : a defence*. Oxford: Clarendon.

Shepski, L. (2008) 'The vanishing argument from queerness', *Australasian journal of philosophy*, 86(3), pp. 371–387. doi:10.1080/00048400802000568.

Silverstein, H. S. (1994). [Untitled]. *Noûs*, 28(1), pp. 122–127. doi:10.2307/2215930.

Sinclair, N. (2006) 'THE MORAL BELIEF PROBLEM', *Ratio (Oxford)*, 19(2), pp. 249–260. doi:10.1111/j.1467-9329.2006.00323.x.

— (2020) *Ethical Subjectivism and Expressivism*. Cambridge: Cambridge University Press.

— (2021) *Practical expressivism : a metaethical theory*. First edition. Oxford: Oxford University Press.

Sinhababu, N. (2009) 'The Humean Theory of Motivation Reformulated and Defended', *The Philosophical review*, 118(4), pp. 465–500. doi:10.1215/00318108-2009-015.

Slovic, P. (2007) 'If I Look at the Mass I Will Never Act " : Psychic Numbing and Genocide', *Judgment and decision making*, 2(2), pp. 79–95. doi:10.1017/S1930297500000061.

Smallwood, J. and Schooler, J.W. (2015) 'The Science of Mind Wandering : Empirically Navigating the Stream of Consciousness'. doi:10.1146/annurev-psych-010814-015331.

Smith, M., Lewis, D. and Johnston, M. (1989) 'Dispositional Theories of Value', *Supplementary volume - Aristotelian Society*, 63(1), pp. 89–174. doi:10.1093/aristoteliansupp/63.1.89.

Smith, M. (1994) *The moral problem*. Oxford: Blackwell.

Sobel, D. (2016) *From valuing to value : towards a defense of subjectivism*. Oxford: Oxford University Press.

'Somalia' (2023) *Britannica Academic*. Retrieved 30 March 2023, from https://academic-eb-com.nottingham.idm.oclc.org/levels/collegiate/article/Somalia/108375.

Steckel, R.H. (2008) 'Biological Measures of the Standard of Living', *The Journal of economic perspectives*, 22(1), pp. 129–152. doi:10.1257/jep.22.1.129.

Stell, A.J. and Farsides, T. (2016) 'Brief loving-kindness meditation reduces racial bias, mediated by positive other-regarding emotions', *Motivation and emotion*, 40(1), pp. 140–147. doi:10.1007/s11031-015-9514-x.

Strandberg, C. (2013) 'An Internalist Dilemma – and an Externalist Solution', *Journal of moral philosophy*, 10(1), pp. 25–51. doi:10.1163/174552412X625754.

Street, S. (2006) 'A Darwinian Dilemma for Realist Theories of Value', *Philosophical studies*, 127(1), pp. 109–166. doi:10.1007/s11098-005-1726-6.

— (2008) 'Constructivism about Reasons', in Shafer-Landau, R. (ed.) *Oxford Studies in Metaethics: Volume 3*. Oxford: Oxford University Press, pp. 207–245.

— (2009) 'In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters', *Noûs (Bloomington, Indiana)*, pp. 273–298.

— (2010) 'What is Constructivism in Ethics and Metaethics?', *Philosophy compass*, 5(5), pp. 363–384. doi:10.1111/j.1747-9991.2009.00280.x.

— (2012) 'Coming to Terms with Contingency: Humean Constructivism about Practical Reason', in Lenman, J. and Shemmer, Y. (eds) *Constructivism in Practical Philosophy*. Oxford: Oxford University Press, pp. 41–59.

Struhl, K.J. (2020) 'WHAT KIND OF AN ILLUSION IS THE ILLUSION OF SELF', *Comparative philosophy*, 11(2), pp. 113–139. doi:10.31979/2151-6014(2020).110208.

Svavarsdottir, S. (1999) 'Moral Cognitivism and Motivation', *The Philosophical review*, 108(2), pp. 161–219. doi:10.2307/2998300.

Szekeres, R.A. and Wertheim, E.H. (2015) 'Evaluation of Vipassana Meditation Course Effects on Subjective Stress, Well-being, Self-kindness and Mindfulness in a Community Sample: Post-course and 6-month Outcomes', *Stress and health*, 31(5), pp. 373–381. doi:10.1002/smi.2562.

The Acts 9: 1–22, Holy Bible. Authorized King James Version with Apocrypha.

'The Tale of Genji' (2023) *Britannica Academic*. Retrieved 8 June 2023, from https://academic-eb-com.nottingham.idm.oclc.org/levels/collegiate/article/The-Tale-of-Genji/71039.

Thompson, B. (2006) 'Moral Value, Response-Dependence, and Rigid Designation', *Canadian journal of philosophy*, 36(1), pp. 71–94. doi:10.1353/cjp.2006.0005.

Thoreau, H. D. (2018) *Walden and civil disobedience*. London: William Collins.

Tolstoy, L. (2010). *War and peace*, translated by Louise and Aylmer Maude. Oxford: Oxford University Press.

— (2014). *Anna Karenina*, translated by Rosamund Bartlett. Oxford: Oxford University Press.

Turiel, E. (2002) *The culture of morality : social development and social opposition*. New York: Cambridge University Press.

Velleman, J.D. (2000) *The possibility of practical reason*. Oxford: Clarendon Press.

Wells, H.G. (2005) *The island of Doctor Moreau*. London: Penguin Books.

Wilkenfeld, D.A. and Lombrozo, T. (2015) 'Inference to the Best Explanation (IBE) Versus Explaining for the Best Inference (EBI)', *Science & education*, 24(9-10), pp. 1059–1077. doi:10.1007/s11191-015-9784-4.

Williams, B. (2008) 'Internal and External Reasons', in Adler, J. E. and Rips, L. J. (eds) *Reasoning: Studies of Human Inference and Its Foundations*. New York: Cambridge University Press, pp. 60–66.

Williams, T. (2009) *A streetcar named Desire*. London: Penguin Modern Classics.

Wong, D.B. (2006) *Natural moralities : a defense of pluralistic relativism*. Oxford: Oxford University Press.

Wright, R. (1995) *The moral animal : evolutionary psychology and everyday life*. London: Little, Brown.

Yeats, W. B. (1968) *Essays and introductions*. New York: Macmillan.

Zeidan, F. *et al.* (2010) 'Mindfulness meditation improves cognition: Evidence of brief mental training', *Consciousness and cognition*, 19(2), pp. 597–605. doi:10.1016/j.concog.2010.03.014.

Zeidan, F. *et al.* (2011) 'Brain mechanisms supporting the modulation of pain by mindfulness meditation', *The Journal of neuroscience*, 31(14), pp. 5540–5548. doi:10.1523/JNEUROSCI.5791-10.2011.