THE UNIVERSITY
of EDINBURGH

Moral Agency

An Embodied Narrative Approach

R. E. Hardt

PhD Philosophy

The University of Edinburgh

2017

Contents

Acknowledgements

I would like to thank my supervisory team: Tillmann Vierkant, David Ward and Elinor Mason. In particular, as my primary supervisors, I'd like to thank Till for continually redirecting my focus towards clarifying the overarching narrative of my thesis, and Dave for his meticulous attention to detail. It was a great team and I was well nurtured.

This project was made possible through the University of Edinburgh awarding me an online Career Development scholarship. I also feel grateful for the comradeship of my postgraduate peers and for the dynamism of the philosophy of cognitive science community. The philosophical community at Edinburgh University has created a buzz of intellectual inquisitiveness that has sustained my interest and motivation.

My philosophy A level teachers, Christopher Warne and Ian Claussen, entertained not only with a faux French romance but also through their passion for philosophy. Their encouragement and enthusiasm helped initiate this journey.

There are several people whose support was indispensible to this thesis being pursued, especially through various first year trials. My parents and my friends – particularly Evie Highton, Di Yang, Julie Crow & Joseph Dewhurst – kept me bound together.

Finally, I would like to acknowledge the immense privilege that it has been to spend 3.5 years thinking, talking and writing about the things that fascinate me. I've been provided with a desk, a scholarship, and concepts to wrestle with, and it has been captivating.

Moral Agency

An Embodied Narrative Approach

Abstract

In this thesis I propose that emotions and rationality are integrated, and jointly constitute our moral agency. I argue against the influential 'sentimentalist' claim that emotions are the only constituents of the moral reasons for which we act, by showing that emotions are inextricably bound up with our sensory and conceptual capacities. In contrast, I propose we act for moral reasons when we act in light of the narratives we create and understand. Narrative understanding here is the capacity to inhabit a chain of events. It is embodied and action-orientated, and is co-constituted through our emotional, conceptual and sensory capacities.

Moral Agency

An Embodied Narrative Approach

Précis

The aim of philosophy, abstractly formulated, is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term.
Sellars, 1963

My thesis asks: what is the role of emotions in our capacity to work out, and act on, our moral values? My answer is that emotions constitute our narrative understanding, and it is through our narrative understanding that we work out, and act on, what matters to us. Because emotions are embodied and action-orientated, so is our narrative understanding. However, narrative understanding is also constituted through our conceptual and sensory capacities, both of which are integrated with emotions.

In putting forward this theory, I retain an insight of Prinz concerning what emotions are and how they contribute to our moral judgements. However, unlike Prinz, and many others, I reject the dichotomy between deliberative reasoning and emotions, and claim that emotions, in the form of narrative understanding, are constitutive of our moral deliberation.

Chapter 1

Emotions and Moral Judgement: A Critique of Prinz

My first chapter examines the theory put forward by Jesse Prinz (2006 & 2007) that emotions constitute moral judgements. By this he means that emotions, when in a compound state with a representation of their object, are constitutive of moral judgements. However emotions are constitutive of moral judgement when, and only when, they are caused by a sentiment attributable to the agent. For example, the conjunction of my anger with the representation of ISIS may be constitutive of a moral judgement. It is a moral judgement if, for example, my anger is caused by me having general sentiments against authoritarianism and unjust violence. A sentiment, here, is a disposition to feel a certain way that is stored in long-term memory.

That my anger is caused by notions of injustice makes it, for Prinz, a moral emotion. A moral emotion for Prinz, is an embodied appraisal triggered by a 'calibration file' pertaining to moral issues. An embodied appraisal is an embodied representation of what a situations means for an organism. In the case of anger the appraisal may be that something the organism cares about has been harmed or insulted. An emotion is calibrated to an issue when a representation of that issue is the eliciting condition for the emotion. In the example above the calibration file that elicits the embodied appraisal is injustice.

While, for Prinz, most emotions are not normally conceptual, Prinz states that moral emotions constitute moral concepts. For Prinz, a concept is a representation that is intentionally controlled. As such, if moral emotions constitute moral judgements, they should be under the control of an agent. A tension in Prinz's account is that moral emotions are not intentionally *controlled* but *caused* by intentional processes, and thus, while he states that moral emotions constitute concepts, they do not seem to play the same role that Prinz gives to concepts. Concepts, for Prinz, participate in our rational deliberative processes, whereas moral emotions are triggered by deliberative processes. His claim about moral judgements therefore boils down to: embodied appraisals, along with a representation of their object, and when elicited by rational processes, constitute moral judgements.

I will be accepting some of Prinz's premises: first that emotions are embodied appraisals, although I will not be understanding them as representations, and second that the processes underlying moral judgements necessarily have some motivational force. That is, I'll be accepting his moral internalism.

However, by the end of my first chapter it is clear that there is some ambiguities in Prinz's account of moral judgements that result from his simultaneous commitments to sentimentalism, his belief that moral emotions constitute moral concepts, and his theory that concepts are under intentional control. The combination of these three commitments, I argue, is why Prinz explicitly states that moral emotions constitute concepts, while describing their

role in our mental processes differently from the role he normally assigns concepts.

My aim throughout the rest of the thesis is to give an alternative account of the processes underlying moral judgements that contests Prinz's view that moral judgements fall outside of our deliberative capacities. This framework makes our narrative understanding constitutive of the moral sense through which we make judgements. Such a framework is a shift in perspective in that the meaning and relationship between terms such as 'agency' 'judgement' and 'concept' are reconceived. And, on this framework, moral judgements are the result of capacities that are jointly emotional and deliberative, rather than just emotional.

Chapter 2

The Mystery of the Missing Agent

Apparently countering Prinz's story is an observation made by Gerrans & Kennett (2010) that a moral judgement exists for an agent that can act for reasons. To act for reasons, they submit, it is necessary to be a creature that can understand itself. For, following Velleman (1992), to have reasons is to act in light of your self-understanding, that is, what is most coherent given your values, attitudes and aims. Human self-understanding, on Gerrans & Kennett's (G&K's) picture, takes places in an autobiographical context, which involves both explicit self-knowledge and the experience of existing through time provided by mental time travel. Mental time travel (MTT) requires not only that we know that, in fact, our past was like this, and our future might be like that, but also that we subjectively inhabit our past and possible future.

In this chapter, I explore exactly where the point of contestation is between G&K's account and Prinz's. It is not quite where G&K believe it to be.

G&K's focus in on how other theories of moral judgement leave out the diachronic capacities, such as an autobiographical knowledge and MTT, that are constitutive of an agent. These capacities are therefore required to make moral judgements, because it is agents that make judgements. For G&K, Prinz's sentimentalism fails because it leaves out these diachronic capacities. This

emphasis, I argue, gives Prinz an easy route out. Prinz can accept that diachronic capacities are important, but because he makes use of the distinction between cause and constitution, he can simply argue that diachronic capacities are important causes of the emotions that constitute moral judgements, but not themselves constituents of moral judgements.

However, lurking underneath these less decisive arguments lies, I argue, a more fundamental disagreement about what conceptual framework best allows us to make sense of agency. We can see this if we return to the notion of agency, as a creature that acts for reasons. Here, I apply an argument of McDowell's to Prinz's schema. Both agree that a creature, to be an agent, must act on processes related to deliberative reasoning. Further, Prinz, like McDowell, agrees that it is concepts that participate in such reasoning. However, McDowell argues that to call an attitude a judgement, it must itself be conceptual, rather than just caused by concepts. As McDowell puts it, a judgement must fall in 'the space of reasons'. Yet, Prinz argues that emotions are not involved in deliberation, and so I suggest that his account of moral judgements, when we take up an alternative viewpoint, fails to be an account of judgements. Seen through McDowell's framework, for as long as Prinz is committed to arguing that our deliberative capacities are only a cause of, rather than constitutive of, our judgements, it turns out that G&K are right to suggest that Prinz does not give us an account of agency or judgements.

This argument throws into relief two very different frameworks for understanding agency, concepts, and judging. While for Prinz, judging can be *caused* by an agent's deliberative capacities, for McDowell, judging is *constituted* by an agent's deliberative capacities. While I will not be proving that McDowell's theory is preferable, this discussion does enable us to recognise that Prinz's argument only works because of fundamental, and debatable, commitments concerning how best to understand these notions and their relationship.

Having laid out the contributions of G&K on the moral judgement debate, I move on to the aims and methodology that will characterise the rest of this thesis. I aim, like G&K and Prinz, to not only build on particular philosophical frameworks about what characterises moral agency and moral judgements, but

also to contribute to a naturalistic framework concerning what psychological and neurological processes enable agency and judgements. Like G&K and Velleman, this view of agency and judgement will focus on our capacity for self-understanding, but like Prinz, emotions will be central to this account.

My methodology is to co-ordinate, and show the convergence between, various lines of enquiry: conceptual, phenomenological and empirical. Prinz is not proven to be wrong, instead, what I construct is a plausible alternative that makes sense of various lines of thought and types of evidence, some of which is also a problem for Prinz's account. The unification of seemingly diverse existing theories under the theory of narrative moral agency, the philosophically informative novel contributions to it, and the accounting of various empirical results using this theory, are proposed as reasons in favour of the theory of narrative moral agency.

Chapter 3

Emotion in Narrative Understanding and Mental Time Travel

We leave Prinz for the next two chapters to start developing a positive proposal of agency. Prinz will be returned to when some new theoretic tools are in place. In this chapter, I pick up on G&K's claim that MTT is crucial for being able to make and act on responsible decisions. I bracket agency temporarily, given its philosophical weight. In this chapter, I argue that we can give an alternative, but related, explanation for the empirical evidence G&K use to support their claim about the importance of MTT to making and acting on decisions. I develop the view that it is narrative understanding, rather than MTT, that is central to these capacities, and which can explain the problems people with ventromedial prefrontal cortex (vmPFC) damage have with acting responsibly.

Narrative understanding is proposed as a broader category, of which MTT is a specific type. MTT is telling and understanding stories explicitly about oneself. Crucially, not all the narratives we understand are explicitly about ourselves. Particularly, unlike G&K, my account of narrative understanding does not give particular importance to our capacity to represent ourselves as psychological beings, a capacity known in the rest of this thesis as

metarepresentation. Narrative understanding may only implicitly involve the self, in the sense that narrative understanding is a perspectival and embodied activity. Watching the news, I may understand events narratively, as the newsreader takes me through an emotional sequence of events.

To develop a theory of narrative understanding I draw on the work of Velleman and Goldie. Narrative understanding, for Velleman, consists of understanding events emotionally. Since emotions are an embodied sense of situations, narrative understanding is the understanding of a sequence of events viscerally and kinaesthetically. We understand narratives because the sequence of emotions in them has cadence. Narrative follows patterns that are familiar to us, because these patterns mimic patterns that are common in everyday life, and because there is a sort of logic to what emotions are likely to follow from other emotions. For example, it makes sense for grief, as the feeling that something we have love has been lost, to follow from love.

Using Goldie's insight that narrative understanding is characterised by a perspective on events, but might also contain perspectives internal to events, I relate narrative understanding back to MTT. Narrative understanding is the general capacity to have an emotional and diachronic understanding of events that are not currently before us, regardless of whether they are about ourselves or not. As a perspective co-emerges with emotion, all narrative understanding elicits an emotional cadence in us as onlookers to the event. We can also understand ourselves as characters in a story when the emotional cadence we experience co-emerges with the perspective of a character in the narrative. In cases of MTT, this character is us. To have an embodied perspective on events, means to have an embodied sense of what those events mean to us, including a sense of what actions a situation affords.

I argue that we also have scientific reasons for believing that emotions are important for our sense of perspective. Work in neuroscience indicates that regions of the brain associated with emotion are also associated with our sense of self. In particular, engaging in narrative and autobiographic memories increases the activity in emotion-associated brain regions. This is what we would expect, given the theory above. If emotions are perspectival, in the sense

that they are an embodied sense of our selves engaging in a certain situation, then we can make sense of this evidence: brain regions associated with emotion are involved in narrative activity and autobiographical memory because emotions co-emerge with an embodied sense of being engaged in a situation.

Returning to G&K's claim that the vmPFC is involved in acting responsibly through its contribution to MTT, I argue that there is preliminary empirical support for accepting that the vmPFC is involved in a broader function than that of enabling MTT. It appears to be activated in tasks that involve taking a perspective, whether one's own or another's, on a complex temporal sequence or spatial situation. This appears more akin to an involvement in enabling a general process such as narrative understanding rather than MTT. That the vmPFC contributes to our ability to act responsibly through its contribution to narrative understanding is therefore a plausible alternative to G&K's theory that the vmPFC contributes to acting responsibly through its contribution to enabling MTT.

Chapter 4

Narrative Agency

Chapter 4 looks at the relationship between narrative understanding, mental time travel, metarepresentation and agency. I first argue that, contrary to Velleman, our metarepresentational self-understanding – what he calls causal-psychological understanding – depends on, and is continuous with, non-metarepresentational narrative understanding. I argue further that minimal (i.e. non-metarepresentational) narrative understanding is sufficient for the type of unified self-understanding that is constitutive of agency. Finally, I argue that a minimal narrative understanding account of agency can account for Strawson's critique of narrative views of self, and is better at accounting for anthropological data than metarepresentational accounts.

Both G&K and Velleman make metarepresentation central to agency, but it is unclear how this relates to narrative understanding. Velleman (2007) claims that these capacities are distinct. I argue that they are interconnected.

The diachronic perspective that emerges with narrative understanding explains how we are able to engage in causal-psychological explanation.

Then I propose that minimal narrative understanding is sufficient for fulfilling the functional role needed for agency that is normally attributed to causal-psychological understanding. When we understand something narratively, we engage in the formation of a coherent world-view that is always also the formation of a coherent self-understanding. This is because narrative understanding is inherently perspectival and embodied – that is, it contains a prereflective sense of self – and because it involves conceptual capacities that enable us to draw coherent inference. Thus, even while we are directed towards the world, rather than our own psychologies, our understanding of the world contains a sense of who we are and what we care about. Like causal-psychological accounts of agency, narrative self-understanding enables agency by unifying us. When we have a relatively unified self-understanding, it is then possible to act in a way that is consistent with who we are. This counts as a reason for action.

Once we understand how narrative understanding is sufficient for playing this role in agency, we can acknowledge Strawson's criticisms of narrative theories of the self. We can agree with him that metarepresentation is not something that all people engage in, but we can also see why this is not a problem for narrative theories of agency. Narrative theories of agency do not require that narrative understanding is metarepresentational understanding. Furthermore, if we want narrative theories of agency to count in the right people as agents, they better not require that agency requires metarepresentational understanding, since there are some groups of people who don't generally metarepresent, and yet it seems wrong to suggest that this compromises their agency.

Chapter 5

Narrative Moral Agency

The account of agency I develop in chapter 4 feeds into my account of moral agency in this chapter. Here, I return to issues encountered in the first and

second chapter: the relationship between emotion and deliberation. And in doing this I return to my debate with Prinz. I develop an alternative model of moral judgements that relies on McDowell's framework. This model develops the view of narrative agency I have already begun, incorporates theoretical and phenomenological insights from Taylor, and fits with empirical evidence concerning narrative understanding and concepts. Throughout this chapter, an alternative account of moral judgements, which makes Prinz's distinction between deliberation and emotion incomprehensible, will come into greater focus. The theoretic considerations and empirical evidence that this account makes sense of, act as reasons to take it seriously as an alternative to Prinz's theory. In all, while we have seen Prinz equivocate on whether emotions are conceptual, in this chapter I will argue that we should understand them as conceptual in the sense that they are involved in our capacity to stand back and reflect on what we believe and why.

Taylor's explanation of what he calls 'strong evaluations', which is our sense of what is moral, situates them as both emotional and constituted through our capacity for deliberation. In particular, strong evaluations are our sense of what is of qualitatively higher and lower worth. They are constituted by a narrative network in the sense that they encompass an interconnected, and embodied, conceptual network that emerges with a particular moral perspective.

While strong evaluations are affective and embodied they can also be articulated, clarified and called into question. Further, they are conceptual in the sense that they incorporate the moral language we have available. Because we can be asked to justify our strong evaluations, and because they incorporate our moral language, Taylor's strong evaluations are both emotional and placed in the space of reasons. As such they fall within the same framework as that explained by McDowell.

This contrasts with Prinz's framework, where emotions are caused by deliberative capacities but are not themselves involved in deliberation. Instead, on a narrative view, emotions have the potential to be involved in deliberation, and in this sense they are fully conceptual. Emotions are a conceptual capacity

in the sense that they participate in what I call a 'recombinant system'. That is, our emotions participate in sequences that can be combined in various ways. Further, our emotions also take part in the type of recombinant system that we can use to deliberate. We can articulate our emotions to justify our beliefs and actions. This type of recombinant system I call 'language'.

As we saw in chapter 3, our narrative understanding and our perspective emerge together, and in this chapter I argue that, when our narrative understanding engages a sense of what is higher and lower worth, it co-emerges with a relatively unified moral perspective. Since it is through our moral perspective that we make moral judgements, and our moral perspective is conceptual in that it exists in the space of reasons, moral judgement are not only emotional, but also conceptual. Specifically emotions are conceptual in the sense that they are capable of being involved in deliberation.

The result of this difference between Prinz's and my theory is two very different stories about how we experience moral reasoning in life. On Prinz's framework we engage in thinking that culminates in an emotion. On mine, we engage in understanding narratives, which is jointly affective and conceptual throughout the process.

Finally, I argue that there is empirical support for a narrative account of moral agency. First there is support from neuroimagining studies on moral judgements and narrative comprehension that shows that there is a correlation between understanding stories and the activation of the medial prefrontal cortex (mPFC), of which the vmPFC is a part. This study also shows a correlation between activation of the mPFC and moral judgements. This fits with my empirical proposal that the neurological enabling conditions for moral judgements involves the mPFC, and that this region enables narrative understanding. Second, there is evidence that abstract concepts – which I suggest is what moral concepts are – are affective. Since such abstract concepts are abstract words, and therefore concepts we use to deliberate, we ought not make the distinction between deliberation and emotion that Prinz makes.

Chapter 6

The Interdependence of Sensory and Emotional Experience

In my final chapter I return to developing an account of how narrative understanding, and therefore agency, is essentially embodied. To do this, I address the topic of sensory experience, something that G&K make central to their account of moral agency by their inclusion of MTT. Remembering and imagining events appears to be a sensory activity. We visualise a scene, or imagine a tune, for example. So far, in my account of narrative understanding, I have focused on its emotional character. So it does not seem clear that narrative understanding is importantly sensory. In this sense it appears that I am at odds with both Prinz and G&K but for different reasons: Prinz because he thinks that moral judgements are a compound state that includes an emotion and a representation of the emotion's object, and G&K because MTT is sensory. However, this difference is only apparent. While I agree sensory experience is part of agency, I argue that emotions and sensory experience are interdependent rather than independent, as Prinz claims. I argue that my theory is preferable to Prinz's because it accurately describes the phenomenology, it enables us to understand how the phenomenology and empirical evidence are consistent, and it gives us novel insight into cases in clinical psychology. The phenomenological observation that our sense of perspective is constituted through the interdependence of emotional and sensory experience helps us to explain the experience of people with depression and psychosis. And it is this observation that also explains the perspectival nature of narrative understanding.

I argue for the interdependence of sensory and emotional experience first by examining the transcendental argument put forward by Merleau-Ponty. This argument, if it works, means that the interdependence of our experience of body and world is a necessary condition for us to have any experience at all. Using this as a starting point, I characterise the way that the entanglement of sensory and emotional experience are evident in everyday life. This poses some issues for Prinz's arguments for the independence of sensory and emotional experience based on phenomenological observations.

Prinz may claim that the phenomenology can come apart from a neurological theory about the relationship between emotions and sensory experience. However, the current neuroscience is consistent with the phenomenology, and this fit is offered as support for my theory over Prinz's.

Finally, while it may seem that we can have emotional disturbances that are isolated from sensory disturbances, I argue that evidence from clinical psychology shows this isn't the case when these disturbances are global rather than localised. I focus on depression and psychosis and argue that both involve a global disturbance of sensory and emotional experience. Psychosis, in particular, as an extreme condition, also comes with extreme changes in a sense of self and world, supporting the theory that the integration of sensory and emotional experience is involved in our sense of perspective. Prinz's theory of the relationship between emotions and sensory experience cannot explain this phenomenon.

Moreover, this interdependence of sensory and emotional experience explains why MTT does not have a special characteristic that narrative understanding is lacking. Having explained in more detail the connection between sensing, emoting and acting, I flesh out the sense that narrative understanding is involved in acting. Our virtual narrative adventures involve a virtual embodied prereflective self because of the sensori-affective nature of narrative. In turn this means that our virtual actions and their consequences belong to ourselves in the present because there is an embodied, experiential identity between the self in past, present and future. This makes future and past actions available to us. That is, the emotional cadence of our narrative understanding is a sensori-affective cadence. This cadence emerges with a sense of a diachronic perspective, situating our current perspective within a chain of events. Hence our narrative understanding, while it involves understanding of counterfactuals, presents us with possibilities for action in the here and now. Thus, while I agree with Prinz that sensory experience is constitutive of moral judgements, the reasons for this are not ones that his theory of the relationship between emotions and sensory experience can account for.

Conclusion

We therefore end up with an alternative to Prinz's account of moral judgement, that presents judgements as the activity of a robustly embodied, narrative, moral agent. Nonetheless, this account incorporates some aspects of Prinz's theory of emotions, and maintains, like Prinz, that these are necessary for moral judgements. Moral judgements, on this alternative view, are formed through articulating the moral sense that is constituted by our narrative understanding. Our narrative understanding is also a prereflective self-understanding insofar as its sensori-affective character arises with an awareness of how we are related to the world. And because of its self-involving embodied nature, the possibilities for action that arise through imaginative immersion in a counterfactual situation are available to us in the present. Hence, moral narrative understanding enables us to act on our moral sense.

Emotions and Moral Judgements: A Critique of Prinz

There is more wisdom in your body than in your deepest philosophy
Friedrich Nietzsche, 1961

1. Introduction

1.1. Introducing Prinz

It is reason, rather than emotion, that is typically associated with our personhood and our sense of morality. And emotion is often understood as being counter to reason. So it is surprising, that, in answer to the question 'what is the role of emotion in moral judgement?', Prinz (2006, 2007) answers, 'everything: emotional dispositions constitute our moral judgements'.

Specifically, Prinz offers what we call a *sentimentalist* account of moral judgements. A *sentiment*, here, "is a disposition whose occurrent manifestations… are emotions" (2007, p. 84)[1]. Sentimentalists claim that moral judgements are constituted through sentiments. Praise and blame, or judging right from wrong, *are* sentiments of approbation or disapprobation towards actions, events, people, or other features of the world. An articulated "judgement will be an expression of the underlying emotional disposition" (2006, p. 34). However, Prinz thinks moral judgements can exist unarticulated. Moral judgements, for Prinz, are constituted by a compound state of an emotion and the object to which it is directed (2007, p. 96 & 99). So, the moral judgement "pickpocketing is wrong" (*ibid.*, p. 96) is constituted by anger and a representation of pickpocketing. While Prinz's theory is focused on being sentimentalist about our ability to judge something as morally good or bad, as we shall see in this chapter and the next, this doesn't come apart from a theory about how we can be responsible for our actions. That is, for Prinz, our emotions count as judgements partly because we are responsible for them.

Prinz's sentimentalist theory is originally composed of two major theses about the role of emotion in moral judgement: they are necessary and they are

[1] This theory therefore relies on a form of compatibilism first developed by Strawson (1963) where our ability to judge someone's actions as right or wrong depends on what he calls 'reactive feelings and attitudes', such as moral indignation.

sufficient for moral judgements (see his 2006 paper). This chapter will spell out exactly what is meant by this theory. The necessity thesis will not be rejected in this chapter or beyond, for what is striking about Prinz's theory of moral judgements is not that emotion has a role in them, but that Prinz initially sees emotion as their *only* constituent, and later he continues to prioritise them. However, he does accept that psychological processes other than emotions are crucial for moral judgements, in the sense that other psychological processes play an important causal role in the formation of moral judgements. For Prinz, moral emotions are *merited*: they are generally reached through reasoning and, because reasoning is a thing that agents do, moral judgements are attributable to the agent having that emotion. Furthermore, it is *moral* emotions that are sufficient for moral judgements, where for an emotion to be a moral emotion it must be triggered by, what Prinz calls, a calibration file.

However, by the time Prinz has written his 2007 book, *The Emotional Construction of Morals*, Prinz no longer mentions the sufficiency thesis. While he continues to maintain that moral emotions constitute moral judgements, the relationship between concepts, and therefore cognition, and moral judgements, appears inconsistent. While Prinz continues to be noncognitivist about emotion in general, I will argue that his attitude towards moral emotions is ambiguous. He wants to argue that moral emotions constitute concepts, yet does not give them the same functional profile that he uses to characterise concepts. Such an ambiguous claim about the conceptual nature of moral emotions amounts to, under Prinz's own definition, an ambiguity about the cognitive status of moral emotions. This is a precarious balance, one that I avoid in my own positive thesis of emotions and their relation to moral judgements.

Prinz, and the relationship between emotion and concepts, will be the major focus of this chapter and much of the rest of my thesis. This chapter will explain his overall argument for his theory of moral judgements. In particular, I will be describing the features of his theory that I will be accepting, exploring or contesting throughout this thesis. The role of his theory is to act as a backdrop through which my own positive account of moral judgements can be defined, developed and contrasted.

Prinz's theory opens the door for an embodied account of moral judgement: one that escapes the assumption that moral agency and moral judgement can occur through purely abstract thought. That is, thought that consists of disembodied symbol manipulation, where judgement is an amodal process that is contrasted to our acting, sensing and feeling capacities. Instead, the body takes a central role in moral judgement on Prinz's account because he classes emotions as embodied appraisals that involve readiness for action. Additionally, he thinks we may be able to usefully understand emotions in terms of involving action affordances (2004, p. 228). That is, emotions are experienced as making certain actions available to us. I will be taking his theory of emotions as embodied appraisals on board. Similarly, my positive project will propose a theory of moral judgements that supports moral internalism, a thesis that Prinz also adopts. I depart from Prinz by arguing we should break down the distinction between rationality and emotions further than he does, if we are to see emotions as involved in judgement.

1.2. Setting up the wider dialectic

Before proceeding it is worth understanding how this chapter relates to the wider dialectic of the thesis.

Since Prinz leaves it unsettled how we should think about emotions and rationality I will develop a multifaceted answer to the question 'what is the role of emotion in moral judgement?'. This positive account is the focus of the forthcoming chapters. It will incorporate Prinz's core claims about the nature of emotions and develop their implications. A theory of emotions as embodied appraisals will be put in the context of our capacity for narrative understanding. On this account emotions are important for moral judgements, but not through excluding deliberative capacities.

Narrative understanding – as a type of sense-making[2] that only exists through the interdependence of our conceptual capacities, emotion and sensory experience – is a crucial way that emotion is constitutive of moral agency. Similar

[2] By 'sense-making' I mean the capacity of a creature to make meaning out of its interactions with the world.

to Prinz's proposal, narrative understanding is not detached, abstract sense-making, but intrinsically depends on us as sensing, feeling, embodied creatures.

Therefore, my positive proposal for the role of emotions in moral judgements is: they form part of a web of interwoven processes that jointly constitute our capacity for expressing, determining and acting on what matters for us via their role in narrative understanding. Such understanding is essentially body involving. But this positive proposal will not work within the same intellectual tradition that Prinz works within. As we shall see in the next chapter, my background understanding of how we should understand agency is one expressed by McDowell: agents and their judgements must be understood as falling in the space of reasons, which means that judgements involve concepts. Particularly, in contrast to Prinz, this does not mean that emotions are just triggered in ways that we can control. Instead, they enable us to engage in rational activity itself. So what I will present is an alternative that involves reshaping the conceptual landscape, and therefore highlights that Prinz needs his own set of fundamental commitments to make his argument work. Further, this thesis sets out an account of moral judgements that overcomes the tensions and problems in Prinz's account, as well as uniting various philosophical theories and empirical data, and developing our understanding of cognition and action.

In regards to the present chapter there are a couple of things to note. I will not be attempting to disprove Prinz's claims. However, I will be setting out where I think he is confused. In future chapters, I will explicate a convincing theory of moral agency in which moral judgements are understood through the way our conceptual capacities are integrated with our emotional capacities. This theory of moral judgement brings developments in embodied theories of cognition to the debates in moral philosophy about agency and moral judgements. It also develops, clarifies and expands those theories of cognition, by examining the relationship between affect, action and language. When cognition is actively engaged in, I take it to involve explicit, deliberative, reasoning.

So, we have in this thesis two competing accounts, both that make emotions central, of how it is possible for a creature to experience the world morally. One is Prinz's sentimentalism, where emotions constitute moral judgements, when they

are caused by the right rational processes, and the other is a narrative account, which argues that emotions constitute our rationality.

By the end of this chapter, it should be clear that Prinz's sentimentalism is not always consistent concerning what type of control we have over moral emotions, and therefore how we should see them as related to cognition. Further, I hope to have shown that his sentimentalism succeeds only through equivocating over the conceptual status of moral emotions. However, the ultimate destination in this thesis gives us reasons to take a more decisive stand, through an alternative positive account where emotions and cognition are integrated.

2. Prinz's Sentimentalism

Before the debate begins, we first need to understand some of Prinz's main commitments, including his sentimentalist theory and his moral internalism. Both are intimately bound up with his position on emotions and moral judgements, and his internalism is a commitment that I will share. First, I explain Prinz's sentimentalism.

According to Prinz (2006), "to judge that something is wrong is to have a sentiment of disapprobation towards it" (p. 29). For him, a sentiment is the stance a subject has towards some issue. It is constituted by our disposition to have certain emotions. This means that a 'sentiment' does not refer to a single emotion and a particular sentiment towards something can result in many types of occurrent emotion. For example, if I have a sentiment of disapprobation towards dinosaurs, my stance towards them is one of dislike. I express happiness when I hear dinosaurs are extinct, and feel angry when they are brought back to life.

Moral judgements, on this account, are constituted by emotions, which in turn are occurrent manifestations of moral sentiments (Prinz, 2007). Specifically, moral emotions constitute moral judgements. Moral emotions are emotions that are triggered by, what Prinz calls, a calibration file. So, for example, indignation is a

moral emotion because it is the emotion 'anger' triggered by the calibration file 'injustice'.

Such moral emotions are occurrent manifestations of sentiments. Prinz explains that,

> *We can think of the sentiment in long-term memory as a standing belief, and the emotion in working memory as an occurrent belief. Or, to introduce a useful piece of terminology, we can call the sentiment a* moral rule*, and we can call a particular emotional manifestation of that sentiment a* moral judgement. (2007, p. 96, original emphasis.)

So, the judgement 'that is wrong!' in response to thinking about a friend being dishonest with you, is activated when the memory of the event triggers a sentiment of disapprobation towards dishonesty. This sentiment of disapprobation towards dishonesty constitutes a moral rule that 'dishonesty is wrong'. A particular emotion occurs – perhaps anger at perceived injustice – when the sentiment is activated. When this anger is attached to a representation of the moment of dishonesty, and caused by the concept 'injustice', this compound state constitutes the moral judgement 'that friend has treated me unjustly'.

Prinz initially states that emotions are necessary and sufficient for moral judgements. So let us turn to Prinz's reasons for stating this. Prinz's (2006) defends his necessity thesis through the claim that it is a good explanation of empirical observations. First, it seems that parents teach their children moral rules by the use of emotion. For example, by using punishment to cause fear in the child, or causing distress by encouraging the child to empathise with a person they have harmed. Second, in psychopathy, a deficit in negative emotions correlates with anti-social behaviour and a non-typical way of understanding moral concepts. Psychopaths have trouble differentiating moral wrongs from conventions. These examples, Prinz claims, support the thesis that emotions are "necessary for *acquiring* the capacity to make moral judgements" (2006, p. 32).

Prinz (2006) also thinks there are some reasons to think that a creature's current capacity for emotion is necessary for moral judgements. For one, it is hard to conceive of someone believing that something is wrong without them being disposed to feel negatively towards it. Someone may have all the non-emotional

facts about killing, such as all the deontological and utilitarian reasons not to kill, but not feel negatively towards killing. In such a case it seems we would not say that such a person believes killing is morally wrong.

Prinz also argues that, if moral judgements depended on another psychological process, such as rationality or observation, then there would be more similarity between the moral codes of different cultures. However, considering that moral outlooks vary widely between cultures, geographical regions and political groups, it is unlikely that moral judgement relies on these processes. The assumption making this argument work is that rationality and observation are not prone to the same amount of cultural inculcation as emotions.

Like Prinz, I think emotions are crucial for moral judgements. My argument in the rest of this thesis will give us reason to think that emotion is necessary for moral judgements, through explaining how it is constitutive of our narrative understanding and how narrative understanding contributes to moral judgements. That is, I hope to give alternative reasons for why we should think that emotions are necessary for moral judgements.

However, one thing to briefly remark on now, is that the necessity claim is bound up with Prinz's argument that we need to understand moral judgement as intrinsically motivational. As emotions are usefully understood by many emotion theorists as involving action tendencies (e.g. Frijda, 2004), and some think that other types of cognition are motivationally inert (e.g. Roskies, 2003), emotions look, to some, like good candidate for being one of the constituents in our capacity for making moral judgements.

In his original defence of the sufficiency thesis, Prinz (2006) uses a study where participants are hypnotised to feel disgust when they hear the word 'often', a word picked because it is an emotionally neutral word. When participants later read vignettes, those that were hypnotised were more likely to judge vignettes that include the word 'often' as morally wrong compared to participants who had not been hypnotised. This apparently shows the feeling of disgust is sufficient for judging that something is wrong, since manipulating the emotional response is enough to produce a corresponding change in moral judgement.

Prinz (2007) later rejects this argument since, in such a case, the emotion of disgust is not caused by a sentiment. Sentiments are stored in long-term memory and constitute what an agent takes to be a moral rule. In the hypnosis experiment, the causal chain is bypassed such that the emotion is triggered without the retrieval of a moral rule. Because of this, Prinz argues that the hypnotically-produced moral condemnations "do not qualify as legitimate expressions of the subjects' moral attitudes" (2007, p. 96).

The other piece of empirical evidence Prinz (2006, 2007) uses to support his argument that *only* emotions constitute moral judgements are studies that apparently show that people – when asked to make moral judgments in certain circumstances – cannot give reasons. In one study, participants were given a number of vignettes, two of which were designed to test moral intuitions: a story about incest and a story about cannibalism (Haidt, Bjorklund & Murphy, 2000). Both stories were designed so that the normal reasons people use to make moral judgements did not apply. For example, in the incest case, there were two siblings, who had consensual sex once, suffered no psychological damage, kept it a secret so that they suffered no social stigma, and there was no pregnancy as a result. Participants were 'dumbfounded' because the reasons they gave for their judgements could be debunked, and yet they would continue to claim the events were wrong. However, what remained was the participants' emotional reaction of disgust. So, the only process that appears needed for moral judgements in the dumbfounding cases appears to be emotions.

Prinz (2007) argues that these cases show that emotions express basic values. The question 'why?' no longer applies: "When we get down to basic values, passions rule. People say incest and cannibalism are disgusting. Murder is abhorrent. Stealing is unconscionable" (2007, p. 32).

However, a switch occurs in Prinz's thinking between 2006 to 2007, which has ramifications for the sense in which we understand his theory as sentimentalist. While in 2006, Prinz argues that emotions are sufficient for moral judgements, in 2007 this claim changes subtly. Here he drops the sufficiency claim, and instead argues that moral emotions constitute moral judgements. This may not sound like

a huge change. But it matters for whether we think of Prinz as a cognitivist about moral judgements or not. Because, as we shall see, moral emotions, for Prinz, are conceptual, while emotions, in general, are not. This makes emotions look like they should be counted as not just reasons for beliefs, but rational in the sense that they can be part of our deliberative processes.

This raises questions for how we ought to think of Prinz in relation to sentimentalism. If emotions are part of rational, in the sense of deliberative, thinking processes, then he seems to be as much a rationalist about moral judgements as a sentimentalist. Furthermore, the sense in which emotions are conceptual is ambiguous in Prinz. Normally, for Prinz, concepts are capable of being under intentional control. However, as we shall see, Prinz's moral emotions don't always seem to play the role that other concepts do in our cognitive activities. What I turn to next, before getting stuck into this debate, is another component of Prinz's theory: his moral internalism. Moral internalism is a commitment that many of the theorists throughout this thesis share, so it will be taken as a ground to build on, rather than one of the premises to debate.

3. Moral Internalism

Moral internalism is the view that moral judgements provide motivation to act. It is a conceptual claim: it is part of the concept of moral judgement that it can motivate action (Döring, 2007). Moral externalists, on the other hand, state that moral judgements are rational judgements, and whether they motivate or not depends on processes that are contingently related to moral judgements. For example, an externalist might hold that we act on the judgement that "hitting children is wrong" not because of that judgement, but because we have a motivation external to that judgement not to hit children, such as empathising with children.

Prinz (2006) argues that we have pre-theoretical intuitions that moral judgements motivate and that empirical evidence supports his sentimentalist theory, and therefore internalism. That is, to support internalism he

"intermingle[s] empirical and philosophical results" (p. 30). Because empirical results show, he argues, that emotions constitute moral judgement, and emotions are action guiding, then moral judgements are too. We will see this evidence in detail in the next section.

Additionally, Prinz thinks that if emotions constitute moral judgements and motivate action, then we can explain why psychopaths have difficulty acting morally. If emotions are necessary constituents of moral judgements, then people who have a more limited emotional repertoire will have a corresponding deficit in moral competence.

This appears similar to Smith's (1996) argument, that we do not recognise amoral people as making moral judgements. Psychopaths are (relatively) amoral. They don't appear to be making moral judgements, because we understand judgements as intrinsically motivating. As Prinz puts it,

> *In the real world, psychopaths are as close as we can find to amoralists: when they say that killing is wrong, they have no inclination to refrain from killing. But I think psychopaths are like anthropologists. They report on morality without making moral judgments.* (2006, p. 38.)

This is at odds with the externalist picture, where we can make a moral judgement, but it motivates only with the addition of a desire. Instead we understand the psychopath as being (more) amoral because they have (less) moral understanding, not because they understand things morally but don't have an additional desire to act on what is moral.

In keeping with this line of thought, Prinz notes that our moral judgements tend to reliably co-occur with what we feel impelled to do:

> *Can one sincerely attest that killing is morally wrong without being disposed to have negative emotions towards killing? My intuition here is that such a person would be confused or insincere.* (2006, p. 32)

We can further note that behaviour reliably changes as our judgements change (Smith, 1994). If someone changes their mind on a moral issue, for example from believing that homeless people are bad people that deserve punishment, to believing that homeless people are victims of circumstance, then their behaviour tends to change; for example, they might stop spitting on homeless people and start getting angry at people who do. In these cases, it seems we are motivated by

our judgements. We would question whether someone had made a judgement if it provided no motivation for action.

For Prinz (2006) sentimentalism provides an explanation of how moral judgements motivate us to act. For example, a moral judgement that eating meat is wrong is an expression of disapprobation. Negative emotions inhibit the occurrence of eating meat, typically by "promot[ing] avoidance, ceasing, intervention, withdrawal, and, when anticipated, preventative measures" (p. 36). So sentimentalism explains the conjunct between beliefs about wrongness and feeling motivated to act in certain ways, by the fact that they are both manifestations of our emotional dispositions.

While it may be noted that not all moral judgements do result in action, Prinz argues that all one needs to say here is that we can have a motive to act without acting. The feeling of a hot dinner plate may motivate me to drop it, but if I'm being a waitress at a formal dinner I might be able to override this impulse. Prinz suggests that,

> *A motive provides a reason for action, and a motivation impels us to act. I think all emotions are motives… But emotions are not always motivations. They do not always succeed in impelling us.* (2004, p. 193)

Nonetheless "the somatic component of an emotion *prepares* us for action" (*ibid*., p. 194, emphasis added). In this way, he is a particular type of internalist, one who does not believe judgements necessarily motivate, but, because they are constituted through emotion, they do provide a motive.

We have seen that Prinz puts forward good reasons for why only internalism can make sense of what we mean by "moral judgements". If he is right, we cannot understand someone as making a moral judgement if they are not motivated by that judgement[3].

[3] Even in cases where it looks like moral judgements come away from how we are disposed to act, on closer inspection, this doesn't appear to be the case. Consider when someone's implicit biases are contrary to their explicit judgements of what is good. Implicit biases are attitudes we have that are evident in our automatic and pre-reflective behaviour. They may contrast with our explicit assertions. For example, Jasmine thinks racism is wrong, but has an implicit bias that associates black people with violence. Some of Jasmine's behaviour may be motivated by her implicit bias and not her explicit assertions. However, if we take her assertions to express a moral judgement, then we

However, like Prinz, I have another reason for adopting internalism: internalism is an outcome of the theory I am proposing, not (just) a motivation for it. On the proposal I will offer in the following chapters, I will argue that the best way of explaining our moral sense is as an embodied orientation to the world that is constituted through narrative understanding. When we take a closer look at how we understand the world morally, it is an embodied, affective process that happens to also makes certain actions available to us. So our moral sense, through which we make judgements, prepares us for action. This approach incorporates phenomenological and empirical evidence about moral judgements.

Moral externalism therefore appears to add a superfluous component to our understanding of moral judgements. Moral externalism states that we can make a moral judgement, but for that judgement to motivate, we also require a desire to act on what is good. But no such additional desire is required for an explanation of how we act on what is good if we have a narrative understanding view of moral agency. On this view, our judgements do motivate. So one reason for adopting internalism is that our best explanation for how we make moral judgements is internalist, because moral judgements requires emotions, and it turns out that emotions motivate us.

This may look like I'm begging the question in that I am both assuming and proving moral internalism. However, this is to misconstrue the argument. There is more than one reason to endorse moral internalism: one approach is trying to work out what we mean by "moral judgement", and the other is the result of building a theory of moral judgements that takes into account phenomenological and empirical evidence. My suggestion is both of these approaches lead to the same conclusion.

---

expect at least some of her behaviour to be motivated by an understanding of racism as wrong. Perhaps Jasmine tries to change her implicit biases by attending more to narratives that undercut the typical narratives about blackness. At the least, she might get defensive at the assertion that she might sometimes behave in racist ways. We cannot understand Jasmine's assertion that racism is wrong as a moral judgement if she has never has any motivation to avoid being racist. Some defensiveness on her part, charitably interpreted, is evidence that she is motivated to not be racist and is therefore frustrated at the implications that she does sometimes act in racist ways. If someone claims they believe 'racism is wrong' and displays no evidence that they are motivated to be non-racist, then we would doubt that they do judge racism to be wrong.

Internalism, however, comes in different colours. One can believe that moral judgements are motivating, but do not involve reasons to act, or one can believe that moral judgements are both rational and motivating (Döring, 2007; Gerrans & Kennett 2010). The later we might call a *rationalist internalist* (Gerrans & Kennett, 2010). "Rationalism states that if it is right for an agent to choose a certain action in a given situation, there is necessarily a reason for him to choose that action in the given situation" (Döring, 2007, p. 364).

However, there is some ambiguity about the best way to understand what counts as a reason. Many think of 'reason' as something explicit and conscious, like what we express in inner speech or to others. However, others think 'a reason' may be based on unconscious, automatic or non-inferential processes (see Gigerenzer, 2004, and Vargas, 2013). In the case of Prinz, we might understand emotions as providing reasons because, as we saw above,

> *A motive provides a reason for action, and a motivation is that which impels us to act. I [Prinz] think that all emotions are motives. Being angry provides a reason, ceteris paribus, to act… but emotions are not always motivations. They do not always succeed in impelling us*[4]*.* (2004, p. 193,)

So, being angry, as the appraisal than you are being insulted or demeaned, gives you reason to defend yourself. And since emotions, on this theory, also involve action tendencies, this appraisal prepares one to act, even while it may be the case that we restrain ourselves.

However, in the sense of 'rational' I will be using, Prinz is not a rationalist internalist. As I will argue below, Prinz does not think that emotions are involved in our deliberative capacities. Instead of being involved in these activities, emotions are caused by them.

[4] We may disagree with Prinz's terminology here. In everyday language, we may often talk of something being a motivation even if it may not succeed in impelling us to act. The threat of early death due to cancer may be one motivation for me to give up smoking, but this motivation may not be strong enough for me to give up, or the competing motivation might override it, perhaps the motivation of being coherent with my self-understanding that I'm happy to live fast and die young. While disagreeing with Prinz's terminology, I think his point stands that not all motives end up motivating us to act in line with them.

In the next chapter, however, I will argue that we have to be rational internalists: if we are to explain moral judgements they must open to justification and clarification. In this way, moral judgements are affective but they are also conceptual. As we shall see in chapter 5, our moral concepts are affectively constituted[5].

4. Emotions as embodied appraisal

Now I want to explain what Prinz means by 'emotion'. Like my discussion of moral internalism, this will introduce some commitments that Prinz has that I will build on. However, it will also be used to explain why I think there are shifts and inconsistencies in his account.

Prinz argues that emotions are embodied appraisals. An appraisal, here, is a "representation of an organism/environment relation with respect to well-being" (2004, p. 52). So, fear is the appraisal that you are in danger. And sadness is the appraisal that something valuable has been lost. Crucially, Prinz argues, these appraisals are embodied. Emotions are mental representations of a set of physiological changes in your body that simultaneously signal how the world bears on you, and prepare you for action. For example, fear is a mental representation of an increasing heart rate and a surge of adrenaline, and represents that you are in a harmful situation. It prepares you to flee, fight or freeze. Roughly speaking, this idea of emotions as embodied appraisals is going to continue throughout this thesis. However, it will undergo some tweaks.

---

[5] That I have endorsed moral internalism may make it puzzling that I argue against the sufficiency hypothesis. The worry is that emotions motivate, and concepts are motivationally inert. So, if one is to be a moral internalist, then it better be that the process involved in moral judgements are processes that can motivate.

It is true that emotions seem to fit the job description of 'a psychological process that can motivate'. However, at the most, this implies that emotions are necessary for moral judgements, not that they are sufficient. Considering that I agree that moral judgements are affective, I retain an explanation of how they can motivate. Prinz himself takes the necessity thesis, rather than the sufficiency thesis, to be crucial for a moral internalist stance.

In forming this theory, Prinz seeks to overcome the opposition between appraisal theories of emotion and theories of emotions as bodily feelings. As he sees it, appraisal theories, where emotions represent how the situation is for the organism, are typically cognitivist theories. Cognitivist theories depend on some account of cognition that differentiates it from bodily responses, and hold that emotions are characterised by cognitive processes. Cognitivists about emotions generally argue that emotions are appraisals or judgements of some sort and are therefore *not* characterised best by our physiological state. The assumption, here, is that (representations of) physiological states are not the types of things that make appraisals.

However, while Prinz accepts that emotions are appraisals, he does not accept that they are cognitive. Which is to say, that, in 2004, Prinz sees his theory of emotions as a non-cognitivist theory of emotion. Act of cognition, for Prinz, involves the intentional use of representation, such as when we explicitly reason. Emotions appear to arise in the absence of such reasoning, so cognition is not necessary for an occurrence of an emotion. We often feel fear as a direct result of what we see, without any intervening, explicit, thought. For example, someone can see a mouse in their house, immediately feel scared, shriek and gather their whole body onto a higher surface than the mouse without first thinking "a mouse! How dangerous!". Quite the opposite: if there is any conscious thought, it comes after the fact, in the form of, "a mouse! That shouldn't be scary".

Most emotions, for Prinz (2004), are like this. They are passive reactions to the world rather than something we are able to wilfully control. He admits that emotions may sometimes be conceptual, "as when we plan for the future. One might wilfully imagine being afraid and elated to determine whether the emotional costs of a roller-coaster ride will outweigh the benefits" (2004, p. 50). But "our ability to wilfully generate emotions does not entail that every episode of emotion is conceptual" because "emotions that are caused in us by events in our everyday life are not concepts. They are more like percepts. They are under exogenous control" (*ibid*.).

Furthermore, Prinz (2007) argues that there are neurological reasons for thinking that emotions are not (normally) conceptual and are embodied. LeDoux

(1996) and Morris et al. (1999) present evidence that (some) emotional responses are primarily the result of the amygdala and thalamus, which are argued to be involved only in bodily changes and not conceptual activity. This appears to the case when we respond to pictures of a coiled snake (LeDoux, 1996) and if we respond to the rapid presentation of facial expressions (Morris, et al., 1999). So, Prinz concludes that "emotions can arise without judgments, thoughts, or other cognitive mediators" (2007, p. 57).

Prinz (2007) also thinks there is reason to think that bodily changes are sufficient for emotion. First, he uses evidence from a study by Levenson et al. (1990) that shows that pulling facial expressions associated with a certain emotion causes both a particular pattern of bodily changes and induces a particular emotional state. Since there is no evidence of a conscious judgement occurring, our embodied state here appears to be sufficient for an emotion. Prinz invites us to experience this too, in the wild. We can smile and notice the feeling of happiness, scowl, and experience the feeling of anger; so the emotion occurs in the absence of thought and the bodily pattern appears sufficient for the feeling. Additionally, Prinz questions the claim that there are ever cases where an emotion is felt in the absence of bodily sensations. When we experience an emotion, it appears that some bodily sensations are always present.

Prinz takes himself to have shown that felt bodily changes and the experience an emotion normally co-occur, by the examples above. So now he turns to particular cases where it might seem like there is an emotion without an embodied component. We shall see that he argues that both these possible exceptions are unconvincing, so he argues that the cognitivist has failed to provide examples of where emotions are experienced in the absence of a particular embodied state. Therefore, the idea that our embodied states are necessary for emotions remains plausible.

The first counter-example cognitivists present are cases where there is a sentiment, but not an *occurrent* emotion. For example, one can have a spider phobia without being in constant fear of spiders. Nonetheless, Prinz responds that one cannot have a spider phobia without the disposition to have, say, butterflies in

the stomach and the rushing, tingling, feeling of adrenaline to the limbs when in the presence of a spider.

The second counter-example the cognitivist presents is the apparently more cognitive emotions such as aesthetic appreciation. For Prinz (2007), even what we think of as more soulful sentiments, such as aesthetic appreciation, do not occur without bodily changes such as shivers down the spine, or an expansive feeling in our chest. Evidence from neuroimaging studies appears to support this: they show that brain areas involved with bodily regulation are active when people view art (Kawabata & Zeki, 2004; Vartanian & Goel, 2004). So far, there are no good reasons for thinking that that bodily changes aren't necessary for emotions.

However, convincing someone that bodily changes are necessary and sufficient for emotions does not show that emotions are embodied *appraisals*. So why does Prinz hold that they are and how can an appraisal theory be consistent with his theory that emotions are a pattern of bodily change?

Prinz holds there can be a difference between what a mental state registers, and what it represents. It is true, for him, that emotions occur when bodily changes are registered. Yet Prinz thinks that for a mental state, like emotion, to represent something, it must have the function of detecting that thing. If emotions represented bodily changes then they would have the function of detecting bodily changes. Furthermore, he thinks that we should also assume that the function of a mental state will have come into being via natural selection, and will therefore tend to enhance the survival of a creature. For Prinz, this creates a problem if we believe that emotions represent bodily changes, as the survival advantage of emotions seems to come from their integration with our decision-making capacities and our behaviour responses. 'Representation of a bodily state' doesn't seem to fit with this criterion. It is not clear why "we should flee when our heart races" (2004, p. 59). And, it is not clear how decision-making would be helped through anticipating non-evaluative bodily changes:

> *Suppose I do not know whether a certain course of action will make my blood vessels dilate or constrict. Does my ignorance lead me to recklessness? If so, it is not clear why.* (ibid.)

Since bodily changes, as long as they have no evaluative character, do not seem to be adaptive, it doesn't seem right to Prinz that they are what emotions represent.

Emotions also seem to be triggered by our perception of our situation. We feel scared when it seems to us our situation is dangerous, sad when we feel we have lost something valuable. We may, as Prinz points out, understand different things as dangerous or valuable. But our emotions are reliably "elicited by things as they relate to us. This suggests that emotions represent relations between external states and our selves. The represent organism-environment relations." (2004, p. 60). Note that it makes sense to see how the understanding of something as dangerous to oneself is involved in prompting one to get away from it now, and make decisions about how best to avoid it in the long-term. So, as Prinz understands it, the theory that emotions represent appraisals of organism-environment relations fulfils both criteria of what it counts for a mental state to represent: it has the function of detecting something (an organism-environment relation), and we can understand how that function is adaptive (it helpfully informs our current and long-term behaviour). Prinz concludes that, "emotions represent changes in organism-environment relations by tracking changes in the body" (2004, p. 78).

I find the direction of this argument broadly convincing. While I have doubts about understanding emotions as representations[6], the observation that emotions systematically co-vary with our situation, and our, often implicit, understanding of it, still holds. So does the observation that their role in our intentions and decisions remains opaque if they do not have content or express meaning. Similarly, the evidence that emotions are embodied seems empirically

[6] This will be expanded on in chapter 5. Briefly, however, it is because the word 'representation' suggests that, for there to be meaning involved in a mental activity, we must make an inference between it and what it is about. For example, a thermometer represents the temperature by us making an inference from the number on it, to what the number is about (Taylor, 1983). While Prinz could respond that we needn't, and he doesn't, mean that by 'represent', there is still a debate about whether more minimal uses of these term is misleading. In any case, the jury is still out on this long-standing debate, and it cannot be settled in this thesis.

and phenomenologically convincing[7]. And, as we shall see in chapters 3, 4, 5 and 6, if we use this model of emotions we can make sense of a whole host of observations about agency.

However, while I will be accepting that emotions are embodied appraisal, I will not be accepting that emotions, in general, are non-conceptual. So, before moving on, I want to challenge the way Prinz's uses the neurological literature to argue that emotions are generally non-conceptual.

While it may be that some studies have shown a small and select neural underpinning for some emotional responses, recent meta-analyses of the literature provide a different picture. It appears that the enabling mechanisms of emotion are widespread and involve areas of the brain that enable cognition.

Many of these areas associated with cognition overlap with Prinz's understanding of cognition. For example, the studies include the activation of brain areas that are involved in cognitive control (Pessoa, 2008) and language (Lindquist et al., 2012). Both of these capacities seem to be included in Prinz's understanding of cognition, given that cognition is understood to depend on our capacity to control representation. Since language-use can be understood as manipulating representations, it fits Prinz's definition quite well. Nonetheless, I note that it is tricky to compare the capacities that some neuropsychologist distinguish to Prinz's definition, given how the former are often orthogonal to Prinz's categories.

Pessoa (2008) argues that the best way to understand the neuroscientific literature is that "complex cognitive-emotional behaviours have their basis in dynamic coalitions of networks of brain areas, none of which should be conceptualised as specifically affective or cognitive" (p. 148). His broad argument is that areas of the brain generally thought to be involved in cognition tend to also be involved in emotion and vice versa. In contrast to Prinz's assertion, the amygdala, along brain regions associated with emotion, "might function as important connectivity hubs" (p. 152). Specifically, the amygdala is highly connected to areas of the brain typically thought to be involved in cognition.

[7] For more evidence, see Colombetti (2014).

Whether we should see it as enabling processes that are completely distinct from cognition, when it appears to generally be involved in cognition, therefore is debatable.

Similarly, Lindquist et al. (2012) found that discrete emotions[8] were associated with widespread activation in brain, including areas associated with conceptualisation, language, and executive attention. All of these are cortical areas, rather than the thalamus and amygdala.

We will see in the rest of this thesis a more detailed account of why we should see emotion and cognition as integrated. For now, all I want to suggest is that Prinz's use of empirical evidence to support his claim that emotions are not normally conceptual is highly contestable. Furthermore, as we shall see below, Prinz's interest in arguing that emotions are generally not conceptual is not clearly carried over to his idea of moral emotions.

Prinz's theory of emotions as embodied appraisals should be viewed in a general context where the conceptual dichotomy between sense-making and embodied activity is being broken down. Various emotion theorists have also moved in this direction in recent (and not so recent[9]) years. Notably, within the field of the philosophy of cognitive science, Colombetti (2007 & 2014), Hutto (2012), Ratcliffe (2005), and Slaby & Stephan (2008), among others, have put forward a similar view. Colombetti (2007) argues that "that the experience of evaluating one's environment is already affective and corporeal" (p. 544). In her 2014 book, Colombetti expands on this theory. For her, all affective processes are embodied, enacted appraisals, which are fundamental to life and cognition. That is, affective capacities are a necessary and fundamental characteristic of the way that that an

[8] Discrete emotions are those emotions that we can name, such as anger or joy, and that occur for a fixed period of time.

[9] See Ahmed (2004) in the afterword of 'The Cultural Politics of Emotion' where she discusses how feminist theorists, for decades, have been the initiators of 'the affective turn'. From the 80s onward, feminists have been challenging the dualism between capacities associated with mindedness, and those associated with the body, through their discussion of affect and emotion. Similarly, Taylor has been expressing a view along these lines from the late 50s. The beginnings of these views can be seen to stretch further back. Merleau-Ponty's *Phenomenology of Perception* was originally published in 1945, and as we will see in chapter 6, implies a similar attitude to emotion.

organism, in its interaction with the world, makes meaning. Hutto (2012), while against the use of representation language in emotion theories, similarly understands emotions as embodied attitudes. In his exposition of emotions, Ratcliffe (2005) argues that bodily feelings are part of the structure of intentionality. Slaby & Stephan (2008) take emotions to reveal "how things stand with regard to our personal well being or our faring in the world in general" (p. 507) through our felt body. Again, emotional experience 'discloses' or 'makes manifest', rather than 'represents', our relation to our world. Many years before, the philosopher Taylor (1985), also conveys a theory of emotions where they are simultaneously expressions of the 'import' of a situation (i.e. how a situation matters to us) and embodied senses. Like Hutto, Slaby and Stephan, Taylor wants to move away from representation language, for reasons that will be discussed later in the thesis. Further afield, feminist, queer theorist and critical race theorist, Ahmed (2004) argues that emotions are bodily orientations to the world.

All of which is to say that Prinz, while apparently making a very bold and (to some) counter-intuitive statement, finds himself within a movement of thinkers who want to overcome the dichotomy between how a creature apprehends the world and its existence as an embodied thing. Unlike Prinz, and more like the thinkers just overviewed, I will be going further than his proposal however. I will be suggesting not only that we should overturn the dichotomy between our embodied activity and our appraisals, but also we should overturn the dichotomy between embodied activity and cognition.

I will be accepting what I understand to be the essence of Prinz's proposal on emotion, which is captured by all of the theorists above[10]. Like a few of those theorists, I will move away from the idea of emotions as representations in chapter 5. While I will continue to agree with Prinz that emotions are embodied, in chapter 2, I will argue that the deep ambiguity that Prinz has in regards to cognition and emotion generates, from one perspective, a dilemma. In chapter 5, I will present my proposal, in response to this dilemma. In chapter 6 I will dispute another aspect of Prinz's (2004) theory of emotions, which is that emotions do not include,

[10] I do not think this essence relies on other facets of Prinz's theory. For instance, I do not share his commitment to the idea that there are basic emotions.

as a constituent, the particular events or objects to which they are directed. So, for example, he argues that if someone feels sadness at the death of a pet, they are in a compound state which includes both sadness, as a representation of something valuable being lost, and (independently of the emotional state) a representation of the dead pet. In contrast, I will argue later that exteroception – sensory information related to the external world – is integral to embodied appraisals. That is, the things emotions are directed at are constituents of the emotion.

Yet I will be affirming Prinz's insight that understanding emotions as embodied appraisals is pivotal to understanding their role in moral judgements. Like Prinz, I will argue that emotions can play this role through the inextricable tie between their evaluative and embodied character. For me, this character is necessary for us to imaginatively grasp the meaning of causes and consequence that aren't currently before us. Unlike Prinz, my theory is not (only) sentimentalist, in the sense that moral judgements are constituted through affective deliberation in the form of making our narrative understanding explicit.

One final thing to note is that there are various distinctions we can make, for example, between 'emotion' 'affect' 'sentiment' and 'mood' (e.g. Prinz, 2004). For example, one can think of a mood as having a general object, and emotion as having a particular one. However, none of these distinctions are relevant to this thesis and I will be using the terms 'affect' and 'emotion' interchangeably. By both, I mean an embodied appraisal.

5. Moral emotions & their relationship to concepts

5.1 Moral emotions

While in 2006 Prinz states that emotions constitute moral judegments, in his book, 'The Emotional Construction of Morals' Prinz (2007) argues that it is 'moral emotions' that are crucial for moral judgements. This brings in a certain amount of ambiguity to the mechanism involved in moral judgement. In particular, this indicates a switch where moral judgements are no longer just emotional, they are also conceptual. However, as I wish to demonstrate, Prinz appears to only

understand emotions as conceptual in some respects, and does not think they participate in our cognitive activity in the same way as other types of concepts do. Specifically, emotions do not seem to participate in deliberation for Prinz.

Here are what Prinz sees as the main moral emotions (2007, chapter 2, section 2.2.), where the equation involves a calibration file and the physiological state associated with a non-moral emotion:

- Indignation = injustice + anger
- Righteous anger = harm against the rights of a person + anger
- Moral disgust = violations of what is seen to be natural/pure + disgust
- Guilt = being responsible for harming someone one cares about + sadness
- Shame = being responsible for violating rules concerning the natural order + embarrassment

In these cases the moral emotion represents a different relation between creature and world than the non-moral emotion it is based on, and yet the calibration file is the cause of this change in the representation, and not a proper part of the emotion. The calibration file is a cause, rather than a constituent of an emotion, like the sun is a cause of sunburn and not a constituent of it (p.68).

It is the addition of a calibration file to an emotion that transforms it into a moral emotion, and a sentiment into a moral sentiment. But, what is a 'calibration file'? Calibration files are a set of "impressions and ideas" (2007, p. 66) that represent a particular concern and trigger a certain emotion. An emotion is calibrated to a set of impressions and ideas when it is triggered by something pertaining to that set. And "moral emotions promote or detect conduct that violates or conforms to a moral rule" (*ibid*., p. 68). For example, Prinz thinks that the moral emotion of indignation occurs when the emotion of anger is calibrated to impressions and ideas of injustice. The calibration file 'injustice' could be set off by thoughts about injustice, or a perception of injustice, say, watching a police person hit a protestor. Such calibration files change the meaning of the emotion, despite the somatic pattern staying the same. So, while anger represents that something you care about, including oneself, has been insulted or threatened, indignation represents that an injustice has occurred.

When the notion of a morally-relevant calibration files is added to Prinz's theory of sentiments, sentiments of dis/approbation now refer not to the disposition towards any types of emotion, but emotions that are moral in the sense that they represent self-blame or praise, or other-blame or praise. A sentiment of disapprobation towards police brutality, for example, results in indignation when such instances are perceived.

Ostensibly, Prinz introduces calibration files to get round what he calls the 'somatic similarity problem'. The somatic similarity problem is that there are presumably fewer distinct patterns bodily patterns of change than there are distinct emotions[11]. That is, what we think of as distinct emotions may share the same pattern of bodily changes. Since, for him, emotions are embodied appraisals because they are somatic patterns that represent a creature-world relationship, and emotions are *only* somatic patterns, Prinz needs to explain how there can be a greater number of creature-world relationships represented than there are somatic patterns. One answer he gives is that emotions can be the result of blends of two or more embodied appraisals (2007, p. 67). Prinz's other proposal is that representing creature-world relations depends not only on the somatic patterns, but also on what caused that pattern i.e. the calibration file (*ibid*.).

To illustrate why this works, Prinz uses the example of a machine that detects smoke, and one that detects carbon. Both are wired to the same alarm. He thinks it would be wrong to say that the sounding of the alarm represents a disjunct, that is, that it represents the presence of "either-fire-or-carbon". Instead, what the alarm represents depends on what has caused it. Similarly, what the word 'bat' represents depends on the context it appears in. If we are talking of a baseball game, it refers to a wooden thing you swing, if the word appears in the context of a full moon and witches, it represents a flying rodent. Similarly, what a somatic pattern represents is determined not just by the pattern, but by what caused this pattern.

[11] I think this might vastly underestimate the complexity of bodily reactions. See James (1884) who understands the body as a complex "sounding board" (p. 191 & p. 202), and Colombetti (2014).

An aspect of Prinz's argument to attend to, given the argument I will make concerning control and emotion, is that his example in no way settles what aspects of the chain of events we should see as constituents of the representation. Is it only the alarm that represents something or is it the machine wired to the alarm plus the alarm? Is it just the word 'bat' or the word 'bat' plus its context? Is it just the somatic pattern, or the somatic pattern plus the calibration file?

Prinz doesn't explain why he takes a causal rather than constitutive approach here. Presumably it is because he has other arguments for thinking that embodied appraisals, in the absence of other capacities, are sufficient for emotions. For example, Prinz argues that emotions appear to be triggered in the absence of the cognitive areas of the brain being activiated, and argues our experience of emotion is an experience of our bodies. So, Prinz's commitment to embodied appraisals being sufficient for emotion, combined with his belief that there is a somatic similarity problem, guide him to his position on calibration files being causes, rather than constituents, of emotions.

Yet, it is not obvious from Prinz's fire alarm thought experiment that we must take this route. I am not currently suggesting that we have good reason to include more into what counts as constituting an embodied appraisal than Prinz does, just that the only criteria we currently have for assessing the fire alarm case are not decisive. That is, it is not clear from the evidence that the content of what Prinz calls 'calibration files' and the affective quality of our embodied experience, are causally rather than constitutively related. Particularly considering the empirical evidence that shows cognition and affect as being enabled by overlapping neural architecture.

What is striking, however, is that calibration files appear to be conceptual in the sense that they are the type of representations that we can manipulate. 'Injustice' and 'human rights', for example, are what we typically take to be the concepts that we use when thinking about morally pertinent situations. So what I want to suggest, and what I want to give more evidence of later, is that Prinz has an interest in keeping calibrations files as causal processes, rather than constitutive, because he wants, to some extent, to reserve a clear difference in kind between deliberative thought and emotions. On this story, calibration files can be part of

cognition, in Prinz's sense, but emotions are not. This is despite thinking that moral emotions are, in some sense, conceptual. After explaining Prinz's argument that moral emotions are concepts, I explain why he does not consistently hold this.

5.2. Moral emotions as concepts

Moral emotions seem to have quite different properties to emotions, in Prinz's view. As we saw above, Prinz thinks that most emotions are not concepts, because we are not generally able to intentionally control our emotions, like we do concepts. However, his view on moral emotions is different, and, I argue, rather ambiguous. In this section I will explain why he thinks moral emotions are conceptual. In the next section I will look at reasons why this claim is more ambiguous than it initially seems.

First, what is Prinz's motivation for claiming that moral sentiments constitute concepts? The basic idea is that an account of judging must involve concepts, in the sense that it is through concepts that we can make judgements. The best evidence of this is in his account of moral Mary, a woman who has no emotions, but who tries to learn everything she can about morality through reading about it. Prinz argues that despites Mary's keen intellectual pursuit of morality, she wouldn't ever acquire the concepts of right and wrong, and therefore would never be able to make a moral judgement (p. 38- 42, 2007). Through thinking about what Mary is capable of, Prinz concludes that while Mary,

> *Can mouth the words "right" and "wrong"… she cannot understand them. This strongly suggests that the concepts right and wrong… are not explicable in terms of the concepts introduced by Kant, Mill, and other normative ethicists*. (ibid., p. 39.)

Furthermore, Prinz argues that if Mary suddenly acquires emotional capacities, she would then posses the concepts right and wrong:

> *The intuitions behind the thought experiment suggest that Mary does not have standard moral concepts until she develops moral emotions. Without moral emotions, she cannot form moral judgements in the ordinary sense.* (ibid., p.42.)

So, for Prinz, an account of moral judgements means an account of moral concepts. And he thinks that emotions constitute those concepts, and that is how moral emotions constitute judgements.

Prinz states that, "we can capture the idea that moral concepts are perceptually based detectors of moral properties by postulating that moral concepts are constituted by sentiments" (2007, p. 94,). Concepts, for Prinz, are perceptually based rather than amodal, so our concept of 'dog' involves "an assembly of perceptual features garnered from our various encounters with dogs" (*ibid*., p. 93,). Since emotions, for Prinz, are perceptions of our bodily state that represent our relation of the world, like other types of perceptions, they can constitute concepts.

Moral emotions constitute concepts when they are representation that reside in "memory or one that has been activated by memory" (2004, p. 46,). This is what sentiments are: they are dispositions, stored in memory, and, when a situation or thought triggers a moral emotion, it is because the situation or thought is triggering a manifestation of this disposition. This is why Prinz argues that, "the standard concepts WRONG is a detector for the property of wrongness that comprises a sentiment that disposes its possessor to experience emotion in the disapprobation range" (2007, p. 94). So, moral emotions involved in feeling disapprobation constitute, when they are stored in long-term memory as a moral sentiment, the concept 'wrong'.

To be clear, it would not be an objection to Prinz to argue that we do not always control our moral emotions, so they cannot be conceptual. Even if moral emotions are not always controlled voluntarily, this in itself would not be a problem for Prinz, who distinguishes between cognition and acts of cognition (2004, p. 46). For something to be a concept, it can often be triggered by the environment. What is important is that it is possible, at other times, to voluntarily use that concept:

> *Consider the case of thoughts that are triggered by perceptual experience. You see a dog and you automatically form the thought that there is a dog in front of you. The thought, and its constituent concepts, does not occur as a result of organismic control. It is a reflex like response to your experience. It qualifies as a thought because the representations it*

> *contains are under organismic control in a* dispositional *sense. You* can *wilfully form thoughts using your dog concept.* (2004, p. 46, original emphasis.)

5.3. Are moral emotions concepts?

5.3.1. Merited emotions

This result, as a reader of Prinz, is somewhat surprising. It seems that, when describing emotions, he has put a lot of effort into arguing that they may be triggered by cognition, but are not themselves, generally, cognitive. Since, for Prinz, cognition involves concepts, in the sense that it involves representations that we could, if we want to, voluntarily manipulate, it is now unclear what the relationship is between moral emotions and cognition. There are several reasons why I think this. First, because he often talks of moral emotions as though they are not part of cognition as we normally understand it. This is evident both in his argument about emotions being merited, and in his use of moral dumbfounding. Second, because it would be hard to make sense of theory as a sentimentalist theory if he did think of emotions as cognitive in the fullest sense. And, if we take his commitment to sentimentalism seriously, then we can understand why he phrases things as he does in relations to merited emotions and dumbfounding. I will expand on all these problems in turn.

For Prinz, emotions can be merited. For him, "moral emotions are not merely caused, they are merited by their causes" (2007, p. 115). Emotions are merited when what they represent applies to the cause of the emotion, and the agent who has the emotion is deemed responsible because they have some control of that emotion. For example, the emotion of 'fear' correctly represents its cause if one really is in danger. In doing so, the feeling of fear fulfils the first criterion needed for a psychological process to be merited. In this sense, emotions are analogous to secondary qualities like colour – our colour perception is working accurately if the experience of blueness is caused by light-waves of a certain frequency or suitable reflectance properties. However, meeting the first criterion isn't sufficient for a process to be merited.

It is the second criterion for a process being merited where Prinz takes the analogy between moral judgements and colour perception to break down. We have control over our emotions in a way that we don't have control over what we see: "in some cases, emotions require a fair amount of deliberation before they arise" (2007, p. 114). For example, one gets indignant at capitalism if one takes time and effort to understand (and then comes to believe) the theory that it is intrinsically exploitative and results in the alienation of the proletariat. While both emotions and colour perception depend on the physical attributes and abilities of the creature, our conscious deliberations are not able to affect our colour perception in the way that they affect our emotions. Even spontaneous emotions can be under control through habituation argues Prinz, for example, we can train ourselves not to be scared when we skydive.

It is important for Prinz to have this distinction because we seem responsible for our moral judgements in a way that we are not responsible for the colours we see. And if moral emotions constitute moral judgements, this means there must be a way that we are responsible for our emotions, because "when we say that something merits an emotion, we imply that the person who failed to have the emotion could be held accountable" (2007, p. 114). For Prinz, having some control over our emotions via our rational capacities explains how it is possible that emotions, and thus moral judgements, can be merited. We have control of our sentiments insofar as "we can deliberate more, acquire more facts, expose ourselves to more experiences, and undergo more training" (*ibid*., p. 115). Because of this, emotions are capacities of an agent, a creature that we hold accountable for making the wrong moral judgement, or not making a moral judgement that they should have. As he puts it, "we cannot hold a colour-blind person accountable for failing to distinguish red from green, but we can hold an emotionally healthy person responsible for being afraid of foreigners or for failing to fear the effects of cigarette smoke" (*ibid.*, p. 114).

Prinz doesn't explain how it is possible that we can change our emotional reactions through deliberation, he just states that we can and do. One possibility is that he thinks that deliberation opens our eyes to how properties of a situation that we have recently recognised are related to old sentiments. If, through

deliberation, one changes from a neoliberal position where capitalism promotes human freedom to a position where it exploits humans and reduces their freedom, then capitalism comes to be seen as an inhumane and oppressive system. If one already had a sentiment of disapprobation towards inhuman and oppressive treatment, then, through deliberation, one comes to feel disapprobation towards capitalist systems.

What I think is important, for our current purposes, about the way that Prinz describes the roles of emotions here, is that they seem to be separated from other types of cognition, and they seem to be caused by other cognitive faculties rather than being voluntarily controlled and integrated with other thoughts. We can see this first in the way that emotions are said to 'arise' through deliberation. Our understanding of deliberation is voluntary, explicit, rational thought where we apply the rules of logic. The idea here, it seems, is that emotions are triggered by our deliberation rather than part of that deliberation, there is some non-emotional deliberation and this causes an emotion. But then, if emotions are not integral to deliberation, but caused by it, then they don't seem to have the same properties that Prinz normally associates with concepts. They do not take part in cognitive acts; instead they are a result of it.

When combined with some other characteristics of Prinz's account of emotions, I think this point becomes clearer. So I move on now to moral dumbfounding, and Prinz's understanding of it.

5.3.2. Dumbfounding

Prinz uses dumbfounding experiments to support his argument that emotions constitute moral judgements. I want to point out that, if he thinks we can control our moral emotions like we control out other conceptual capacities, he falls in to trouble in relation to his explanation of dumbfounding. However, if he doesn't think we can control our moral emotions, his explanation of dumbfounding experiments is consistent with his understanding of moral emotions.

Dumbfounding experiments present people with vignettes, often about taboos but where no harm occurs. For instance, about sibling incest. In the vignettes, the siblings suffer no ill affects, psychological or physical; there is no

pregnancy; and they suffer no stigma because no-one ever finds out. In many cases, participants consider the siblings to have acted immorally.

The dumbfounding claim for the thesis that emotions constitute moral judgements works like this: people make some moral judgement about a particular event; the reasons they provide for their judgement are contradictory to the details of the event; when reminded of this, they fail to change their judgement; therefore it is not explicit reasons that are components of moral judgement. However, their emotional reactions stay the same, therefore it is the emotion, and only the emotion, that explains moral judgements.

Or, more precisely, Prinz thinks that our emotions are basic values that provide reasons but are not based on reason (2007, p. 31-2). If we repeat 'why' to questions about something being wrong, eventually we'll get to 'because it's harmful'. Asked why 'harm' is wrong, and the question seems odd. Suppose we ask, 'why is it wrong to kill people below 5'3" for fun?'[12]. The answer, it seems, is this is harmful and malicious. But ask what is wrong about harm and malice, and the question is baffling. It makes us angry, maybe disgusted, and there is nothing beyond that. This supports, for Prinz, the idea that our basic values provide, but cannot be changed by, reasons, and that they are constituted through affect.

Here, Prinz relies on an observation first made by Hume. Hume (2006) argues that it is only emotion that teaches us of right and wrong. When we reason about moral matters,

> *All the circumstances and relations must be previously known; and the mind, from the contemplation of the whole, feels some new impression of affection or disgust, esteem or contempt, approbation or blame.* (p. 271.)

Sentiments of dis/approbation are reached after considering a system of relations and consequences, but it is the affective sentiment we reach through this that tells us something uniquely moral about the situation. It is emotion and sentiment that

[12] I use this example as a replacement for Prinz's example. Prinz gives the example of the rape of a toddler who will not remember the incident: "Consider the question "why is it wrong to rape a toddler who will never remember the incident?" This is an odd question. It is difficult to answer. It's just wrong to do that." (2007, p. 31). However, this example has difficulties, because a lack of memory does not necessarily co-occur with lack of trauma. Lack of harm, I imagine, is what Prinz thinks this example indicates, but it is not clear that it does.

lie at the root of morality, just as it lies at the root of other decisions about how to live our lives[13]:

> *Ask a man 'why he uses exercise'; he will answer, 'because he desires to keep his health'. If you then enquire, why he desires health, he will readily reply 'because sickness is painful'. If you push your enquiries farther, and desire a reason why he hates pain, it is impossible he ever give any. This is perhaps the ultimate end, and is never referred to any other object.* (p. 273.)

So, one could ask, with the case above, 'why would causing harm be wrong?' And, perhaps the right response is, 'well it just is!'. We feel angry and there is nothing beyond that. Similarly, the emotions elicited in the dumbfounding experiments show that taboos are basic values constituted by emotions. We just find incest disgusting, there is no further reason we can give. Hence we can see Prinz as making a Humean point: our ultimate reasons are constituted by our sentiments[14].

Prinz's argument that emotions constitute basic values appears to be validated by the dumbfounding cases. In these cases, emotions and moral judgements continue to co-vary in cases when deliberative reasons appear to contradict moral judgements. This fits with the notion that emotions provide reasons rather than are part of explicit reasoning practices.

However, there is a puzzle here. In one sense this argument appears to conflict with the account of control over our moral judgements that Prinz gives above, because deliberation seems to have no impact on moral judgements. That is, Prinz seems to imply that no matter what we explicitly recognise, it seems our judgements stay the same. So, while he appears to argue above that we do have control of our moral judgements, now he seems to say the opposite.

---

[13] I wish to leave it open at this stage that morality and decisions about how we live our lives are separate, although this presupposition will be contested later.

[14] Although I want to look at how this argument fits together with the rest of Prinz's theory, we can also question the argument itself. First, it is not just that killing people under 5'3" for fun is harmful, it seems a particularly unjust harm. And there is a tradition, in philosophy, of trying to explain what we mean by harm, when and why harm is bad, what is injustice, what is so disturbing about malice etc. That it is hard to explicate why we have certain intuitions does not make it impossible, or else moral philosophy would not be such a rich domain.

This conflict, as we shall see, is only an appearance. When we see why this is an appearance, another issue in Prinz account becomes more visible, one that I've already discussed above. And that is that emotions, in this account of providing basic values, are not conceptual in the sense that Prinz normally takes a mental process to be conceptual.

First, why would I say that it is an appearance that this account of basic values conflicts with Prinz's account of control over our moral judgements? Well, as seen above, this is because Prinz does not think we have control over our moral judgements in the same way that we have control over the application of logical processes typical in everyday deliberation. While we can understand deliberation, in Prinz's theory, as the intentional manipulation of concepts, Prinz does not think our moral judgements are part of this deliberation, instead they are "elicited" by this deliberation.

We can slot this idea back into the context of dumbfounding and basic values if we understand the argument about merit in a particular way.

We saw that we can understand deliberation playing a part in an emotion being merited by an emotion being elicited by deliberation that enables us to see a circumstance as fitting a certain category. For instance, if I deliberate about capitalism, I might come to see it as unjust. A sense of injustice, like our category of 'incest', can be seen as a basic value. We can see this if we remind ourselves that, for Prinz, moral emotions are formed through particular types of calibration file triggering basic emotions. Moral emotions, therefore, are basic values, which are constrained by what calibration files we have available, and the basic emotions. What our deliberation enables is for us to see some situation as an instance of incest or injustice, for example. I might deliberate some more, and realise that capitalism isn't unjust. In this case, my moral judgement would change in regard to that situation, but the types of moral judgements I have available wouldn't.

Similarly, for Prinz, 'incest' seems to be a calibration file that triggers an emotion. Once we understand a situation as incest, Prinz is arguing, then we

automatically feel an emotion (presumably disgust and/or anger), and that emotion constitutes the judgement of disapprobation[15].

So, either way, moral judgement is inflexible, on Prinz's theory, in the sense that once one understands, through deliberation or perceptions, a situation to be of a certain kind, then the judgement is triggered. Moral judgements are flexible, or controlled, only in the sense that we may not think, or perceive, a situation as being of that kind.

If this is the right way to understand Prinz, then what Prinz says about dumbfounding experiments and emotions being merited are congruent: in both cases, there is a sense that moral judgements are inflexible, in the sense of being automatically triggered by particular perceptions of chains of deliberation. Prinz's analysis of the dumbfounding cases make this clear, because he argues that the moral judgement, which is emotional, is distinct from the deliberative reasoning that leads up to it.

But, if we understand Prinz to be making this claim, then it looks like moral emotions are not conceptual in the way that our deliberative processes are. They are triggered by a calibration file rather than being available for manipulation themselves. However, if moral emotions constitute concepts, then it is unclear why this should be the case. Instead, if Prinz does think that moral sentiments constitute concepts, in the sense that they are under intentional control, then they should be the types of things that participate in deliberation, rather than being triggered by it.

One response to this might be that calibration files are under intentional control, and therefore so are moral emotions. The problem with this is that calibration files are, for Prinz, not a constituent of the emotion, but only its

[15] We might argue with the details of this. Perhaps it would be more accurate to say that categorising a situation as incest seems to automatically trigger a calibration file of "injustice" and/or "harms against the rights of a person" and/or "violations of what is seen to be natural/pure", and this in turn triggers the emotion. Yet this observation would do little to change Prinz's schema concerning how moral emotions work, instead it would be an argument about whether or not to include incest as a basic value itself, or as triggering other basic values.

necessary cause. So, even if calibration files were part of our deliberation, emotions would remain what were triggered by our deliberation.

It appears that Prinz's moral sentiments are both concepts and not concepts under his theory. But why would Prinz create a hybrid of this kind? The answer, I think, is his commitment to sentimentalism, which I will expand on now.

5.3.3. Sentimentalism and deliberation

There is a relatively simple way to explain this ambiguity in Prinz's account as to whether emotions count as concepts or not. The first is Prinz's commitment to concepts being those things that we are able to bring under intentional control. The second is his commitment to sentimentalism: the view that sentiments constitute our moral judgements. The final one is his commitment to the idea that moral sentiments constitute concepts. These three commitments create a problem. Because if moral emotions constitute concepts and moral judgements, and concepts are the types of things we can intentionally control, then it isn't clear that one is still a sentimentalist. If emotions are part of rational deliberation, then one is as much a rationalist about moral judgements than a sentimentalist. After all, if rational deliberation is constitutive of moral judgements, then it looks like one is a rationalist about moral judgements.

It is because Prinz sees himself as a sentimentalist, and not a rationalist, I want to suggest, that he resists giving emotions the full functional role that he normally associates with concepts. By making emotions causally related to, rather than participants in, our deliberations, Prinz can maintain that he is a sentimentalist. So, there is a tension in his account because he has a commitment to sentimentalism, to moral emotions constituting concepts, and to a certain notion of 'concept'. But these three commitments, taken together, create a problem that he tries to resolve by not giving sentiments the role that he normally gives concepts, while simultaneously claiming that they do constitute concepts. This is an unstable position where Prinz simultaneously denies and affirms that emotions take part in our cognitive capacities.

If Prinz, to make his argument coherent, opts to accept rationalism as much as sentimentalism, then that, too would resolve the issue. However, he could not

do that by just claiming that both the deliberative reasoning and the emotion jointly constitute moral judgements. If he does this, he would open up a new worry: it appears on this account that the mental process that provides us with a reason for action is not the same as the process that motivates action.

For if Prinz proposes the above, rational deliberation is the conceptual part of moral judgements that enables our moral judgements to be merited, and emotion is a non-conceptual process that enables us to act. Here Prinz's loses his moral internalism: what provides us with reasons is not what motivates our actions. While moral judgements would be rational and motivate, the part of the judgement that is rational would not be the same part that motivates the action.

Further, it would be unclear why emotions, as the part that motivates but is not part of rationality, could be said to help constitute a judgement. We will see in the next chapter why it is important that if emotions constitute moral judgements they *themselves* are rational, in sense that they can be part of deliberation.

Finally, another route for Prinz out of this problem would be to give up sentimentalism and claim that rational deliberation can motivate actions. This would be a huge departure from his current position. It would also be controversial, and would need arguing for.

So, Prinz falls into additional problems if he responds to my worry about his position being unstable by arguing that rational deliberation and emotions jointly, but separately, co-constitute moral judgement.

A large part of what I will be doing in future chapters is building a positive account of moral judgements that resists this precarious position by giving up any strong claim to sentimentalism, while retaining a central role for emotions. This is a form of sentimentalist rationalism in the sense that emotions and rationality are tightly *integrated*, and therefore co-constitute judgement. In offering this proposal, I sidestep the worries that would arise if Prinz were to claim that rational deliberation and emotion are separate co-constituents of moral judgements. That is, I explain how we could hold that emotions and rationality are integrated.

As we will see in future chapters, a narrative understanding view of moral judgements enables us to see moral judgements as constituted through chains of

affective concepts that exist within a narrative frame. In doing so, I present a coherent form of rational internalism.

6. Conclusion

I have outlined Prinz's sentimentalism and some of the themes connected to it that will be important for the rest of the thesis. I have accepted, to some degree, his theory of emotions as embodied appraisals and his moral internalism. The main point of contention has been his equivocal attitude towards the role of emotions in our deliberative capacities. I have shown that his position is somewhat unstable. An attempt to resolve this instability by making moral judgements comprised of separate parts – deliberation and emotions – leads to additional problems. We will return to this in the next chapter.

One might think that committing to a more rationalist line to resolve this instability might be contrary to moral internalism, emotions being the best candidate for a psychological state that motivates. However, moral internalism seems to more clearly rest on the thesis that emotions are necessary for moral judgements. If one believes that emotions are integrated with our deliberative capacities, then rationalism and moral internalism are not clearly at odds. Further, as shown in chapters 2 & 5, I think we find we have good reasons for adopting, rather than rejecting, rationalism.

Resolving Prinz's unstable position on concepts and emotions by taking a stronger and more coherent stance – and considering some of the arguments proposed by Gerrans & Kennett, Velleman, and Taylor – I propose an account of moral agency which can only be understood through the thorough integration of our conceptual and emotional capacities. Moral judgements, on this story, are the outcome of a multifaceted process that gives a creature reasons to act.

So, while I take on Prinz's theory of emotions as embodied appraisals, I argue that moral judgements are more tightly interlinked with both our deliberative capacities and sensory capacities than he anticipates. What is particularly relevant for this current chapter is that it is not necessarily right that

we see the relation between concept and emotion as a sequential process where the former triggers the latter. Rather, our reasoning is infused with affect, and so the two cannot be separated.

This creates a path out of the unstable situation Prinz finds himself in without running into a new worry: how to avoid being a rationalist that can't explain how judgements motivate, and how to avoid being an internalist that fails to give an account of judgements. It is this pair of dangers that Gerrans & Kennett (2010) are concerned with, and which we turn to next.

The Mystery of the Missing Agent

I've always claimed that the ability to fantasize is the ability to survive, and the ability to fantasize is the ability to grow. Boys and girls of 11, 12, 13, the most important time of their day is especially at night right before they go to sleep, is dreaming themselves into becoming something, into being something. So when you are a child you begin to dream yourself into a shape, and then you run into the future and try to become that shape.
Ray Bradbury, in interview on Day at Night, 1974

1. Introduction

Prinz (2006) has claimed that moral judgements consist of one type of psychological process, an emotional process. On his view, too, moral judgements motivate. On the other side of the debate are theories that moral judgements are constituted through rational deliberative processes that are not intrinsically motivational (e.g. Roskies, 2003). Gerrans & Kennett (2010) take up the task of synthesising these opposing views, by keeping the assumption that moral judgements motivate (i.e. internalism) but finding a way for such motivation to be rational. First, they want to undercut a common assumption, which is that it makes sense to talk of judgements as isolated events. Instead, they propose that judgements are always the activity of the agent that issues them, and thus the task of working out how we can act for moral reasons involves the task of explaining agency in general, rather than judgements in particular. Second, they want to account for agency by arguing that it requires mental time travel (MTT), deliberative reasoning, and emotion, where MTT is the ability to have an experience of subjectively inhabiting one's past and possible future.

However, to understand Gerrans & Kennett's (G&K's) account fully, and to understand how they provide a response to Prinz, we have to understand some of the theoretical backdrop to their argument. To this end, I give an explanation of why the moral psychology and philosophy of J. David Velleman is relevant to their account, something G&K do not expand on themselves. I explain how his theory challenges rival theories of moral judgement by requiring theories of judging and acting to give a role to the agent in the decision-making process. Velleman (2000)

argues that self-understanding is the part of the decision-making process that plays the functional role of the agent.

Returning to Prinz, I ask what it is about G&K's alternative account that is meant to persuade us he has made a mistake. That answer is that it is evident that the act of making a moral judgement is constituted through being an agent that can act for reasons. While Prinz thinks that an emotion can be a reason for action, his idea falls far short of what it would really take to act for reasons, because he doesn't recognise that MTT and other diachronic processes are necessary to have reasons for actions.

However, I argue that G&K have to say more than this. We also have to ask why this has any bearing on Prinz's claim that psychological processes other than emotion make an important causal contribution to moral judgements, but are not constituents of it. Prinz can claim that MTT and deliberation may be important causes of an emotion, and therefore moral judgements, but this is not to say that they are constitutive of them.

What is pivotal here, though, is that some of the commitments Prinz has are up for debate. Particularly, we can debate those commitments that enable him to assume that intentional and epistemological characteristics such as a state being merited can be fully captured using *only* causal language. If, instead, we think that explaining what constitutes *merit* uses a different explanatory framework from causal explanations, then Prinz falls into a dilemma when he tries to integrate causal modes of thinking with his idea that emotions are merited.

On an alternative picture, either Prinz should say that emotions are not merited, because we understand them only through causal discourse, in which case he does lose sight of the agent. Or else, he has to stay true to the task of explaining how it is that we act for reasons, and lose his causal thinking when understanding what constitutes moral judgements. If he does the latter, then his emotions must be conceptual, in the sense that they must be the type of state that can participate in inferential thinking. They must be capable of being justified and of justifying, rather than merely being caused and causing.

Unlike the position we saw Prinz take in the last chapter, emotions are not ambiguously conceptual, but fully conceptual. Because, if we recognise that we

cannot understand what constitutes agency through causal vocabulary, then we recognise that it is a confusion to think of emotions, only mechanistically understood, as being a reason for action. This does not make agency a non-natural, mysterious, phenomenon, for reasons we will see later.

Essentially, this chapter focuses on the same crack in Prinz's argument as the last one, but explains how it is not merely an inconsistency in his theory, but betrays other commitments which we can reject.

The best (perhaps, only) way of understanding G&K's argument as a response to Prinz is that many philosophers take it as central that agency is understood through our participation in the space of reasons (McDowell, 1994). If we take this as fundamental, then the quest is to find a description of our psychological capacities that meets this criterion, and Prinz fails to do this. This reasoning requires us to depart from Prinz's conceptual framework into novel territory. In chapter 5 I will argue that there are also phenomenological and empirical reasons that support this argument[16].

In addition, the explanation of McDowell's framework gives us a way out of the precarious situation Prinz has found himself in at the end of the last chapter. One possibility that comes to the fore with McDowell is that emotions could constitute deliberative reasoning, something I will explain further in chapter 5.

By the end of this discussion, we will have examined and explained much of the basic content of my thesis. What I want to do in the last part of this chapter is explain my aims and the basic shape of my methodology in the upcoming chapters.

We will then have all the pieces in place to begin a detailed explanation of what best explains agency, and how we can depart from Prinz's sentimentalism, while maintaining the importance of emotions to agency. First, I want to explain where G&K understand the problem to lie in contemporary debates about moral judgements.

---

[16] When discussing my methodology later in this chapter, I will explain why empirical reasons are still relevant once we question whether causal explanation exhaust all explanatory needs in regards to agency. Empirical data, as we shall see, provides enabling explanations: how it is possible for a thing like an agent to exist, given that agents are particular type of things. But this last point, explaining what characteristics constitute agency, is not exhausted by the empirical data.

2. Agency

G&K (2010) want to reframe the debate over how we make moral judgements. Many of their peers are misguided, they claim, because they assume that moral judgements are the types of things that depend only on information available in the here and now. One can be a neurosentimentalist, such as Prinz, and take this position, a 'neurosentimentalist' being one who uses neurological and psychological evidence to argue that intuitive and/or tacit emotional processes are necessary and sufficient for making moral judgements. Or one can be a rationalist externalist ('externalist' from now on): one who believes that being able to apply a rule to a particular situation is sufficient for making moral judgements. While the sentimentalist is an internalist, and can explain how judgments motivates, the externalist does not believe that moral judgements are intrinsically motivational.

What we have here is a restating of the two poles that Prinz is caught between at the end of the last chapter if he chooses to argue that emotions and deliberative reasoning jointly, but separately, constitute moral judgements. Both sentimentalists and externalists believe emotions motivate[17], and that they are separate from deliberative reasoning. If we have both these commitments, then an internalist appears wedded to sentimentalism, and a rationalist appears wedded to externalism. An additional problem for the sentimentalist, as we shall see in section 3.4., is that there are problems with arguing that emotions are constitutive of judgements if one also argues that emotions are separate from our capacity to participate in deliberative reasoning.

G&K take a different approach from the one I will present to show that neurosentimentalism and externalism fail to account for moral judgements (although I will also argue that there is something important about their claims). They argue that both tacit processes and the ability to apply rules are synchronic capacities. They claim this because people with amnesia, who are less able to understand themselves as agents extended though time, have the former capacity,

[17] In chapter 6, I will explain why I agree that emotions motivate: I will argue that our capacities to act, to sense the world, and to make embodied appraisals are all caught up with one another.

and some also have the latter capacity. Because people with amnesia are able to have an emotion or engage in logical reasoning, while simultaneously being unaware of themselves as an agent through time, G&K argue that these activities are 'synchronic'.

G&K propose that we should understand moral judgements as an exercise of the capacities of a moral agent and, crucially, such capacities depend on the diachronic characteristics of the moral agent. A diachronic agent, on their account, has two interrelated abilities: they can re-experience their past and imagine their future (which they take to be one capacity, and I will argue the same in chapter 3), and they can think about themselves as an agent through time[18].

> *Planning requires a capacity to imaginatively project oneself into the future; this in turn requires both a sense of oneself as the very same individual who will inhabit that future (autonoetic awareness), and also the kind of detailed self-knowledge that is supported by autobiographical memory.* (p. 601.)

Autonoetic awareness and self-knowledge are important because we become agents through these capacities. They do this, as stated above, by enabling us to make and act on plans and projects.

But why think any of this has something to do with agency? To answer this question, it is helpful to explore the thoughts of some of those they cite, particularly Velleman.

In his 1992 paper, *What Happens when Someone Acts?*, Velleman argues that any description of psychological processes involved in acting must have at least one functional part that corresponds to what we take to be an agentive process. There must be some functional component that *intervenes* by assessing beliefs, desires, in light of some standard, which he initially argues is our desire to act in accordance with reasons. The desire to act on reasons, is "the agent's contribution to the causal order" (p. 479). Having such a functional role in our descriptions of the causal sequence involved in acting allows us to understand the output of the sequence as acting, that is, as the result of agency.

[18] This distinction between experience of oneself through time, and knowledge of oneself through time, will form the basis on much of my discussion in the next two chapters.

Because of this, Velleman argues against the 'belief-desire' explanation for action, arguing that no part of this process has a functional part that is recognisably agentive. Instead, we can understand how the combination of desiring x, and believing that we can have x if we do y, would lead to behaviour y without being able to understand how it is an act of an agent, rather than behaviour of an automaton. In descriptions concerning how the decision-making process of an agent causes agentive acts we want to understand why it is a description of an agent *acting*, something that is done. Instead, standard belief/desire accounts of decision-making can be understood as automatic processes resulting in *behaviour*, something that happens.

In his 2000 paper, *From Self Psychology to Moral Philosophy,* Velleman posits that it is our understanding of our self that we use to assess what we have reason to do. We do not just act indiscriminately on our beliefs and desires since actions are events that are *chosen* according to what makes sense given what type of person we understand ourselves to be, the current situation we are in, and what we understand ourselves to be feeling. We are agents because there is a process that plays the functional role of assessing what we have reasons to do, and we assess what we have reason to do by the above criteria[19]. Importantly, this assessment is driven by a fundamental feature: the desire to be intelligible to ourselves. We can achieve this goal of becoming intelligible to ourselves when we act in ways that make sense, given our characteristics.

There are two broad ways that Velleman thinks we can act on our self-understanding (2007). First, we can build narratives about the events we are engaged with, and decide how to act based on what would make sense, given the line of the narrative. For instance, if Shauntelle has spent the last five-years creating and maintaining a new company, then it makes sense that, even if sales are bad for few months, she will figure out whether there are any particular reasons and try to resolve them, rather than give up straight away. If her company is in a very bad way however, we may wonder why she doesn't give up and start a

[19] This ties in with the analysis of another philosophy G&K cite: Bratman (2000), who argues that our reflectiveness functions to organise us as a singular agent that can act consistently through time.

better company from scratch – it may take less energy to do that, then try and rescue what she has. As we shall see in the next chapter, this is because understanding a narrative involves a particular emotional cadence, and such emotional cadence consists in us making sense of events through what precedes and follows an event. A project we are invested in is understood through its value to us, which becomes apparent only through situating it in a narrative, where what matters is worth pursuing, and so we continue to invest in it despite problems. This brief summary of narrative understanding will be revisited and developed on in chapter 3 and beyond, becoming my main thesis about what makes moral agency possible.

Such narrative self-understanding is largely implicit, we may be aware of what we want and act on it, but we may not think of ourselves as wanting it. Shauntelle might search for reasons she can remedy to keep her company going, but she may not think herself "I really love this company and I'm heavily invested in it. Because of this I want to find a solution". This would count as the second way we can find a rational way to act: we can understand ourselves in causal-psychological terms, which means explicitly understanding ourselves in terms of our psychologies, who we think we are, and finding an action that makes sense, given what type of person we believe ourselves to be. I will return to this distinction in the next chapters, for now what is important is that they are both ways of acting rationally through enabling us to act on self-understanding.

So, to sum up Velleman's thoughts, he asks: how do actions differ from behaviour? His answer: only the former is caused by an agent. What makes something an agent? That it acts for reasons. What motivates this? That a creature wants to be fully rational and coherent in the form of being intelligible to themselves. How do people, as agents, fulfil this aim? They act in light of their self-understanding, which is built either through their narrative abilities or their causal-psychological reasoning, to decide what would make sense for them to do.

Velleman's account of agency is undoubtedly contestable and partial. For instance, he says little about what enables this drive to be self-consistent. It is unclear, for instance, if this is explicitly formulated or implicit in our practices, whether it is innate or learned, individually achieved or socially extended, or some

mixture of these features[20]. However, his basic point about the need for self-understanding and consistency in acting for reasons is widely held (see e.g. Korsgaard, 1989, Bratman, 2000 & Levy, 2014), and we will see in the rest of this thesis that it is a theoretical framework that fits well with phenomenology and empirical evidence. It also forms a coherent network with other theories of agency that elucidate what type of creatures we need to be to be moral creatures (see chapter 5). Finally, to limit my scope, I will unfortunately be circumventing some deep debates in moral philosophy. My perspective, while venturing into the realm of moral philosophy, comes from the philosophy of cognitive science. From now on, I broadly take Velleman's view forming the background of how we should understand agency[21].

This discussion serves to expand on what seems special about agency, which underlies much of G&K's discussion. It shows why it is important that acting in a way that is consistent with our self-understanding is one way of understanding 'acting rationally' and also a way that is often tied to our diachronic capacities, as G&K want to claim. Fundamentally, if we are to *act* in the way that we take people to act, that is by pursuing goals and projects through time, then we need a consistent set of attitudes that justify picking some goals among many.

Importantly for my project, while diachronic agency in the sense of being able to act on long-term plans seems to be important for a lot of human agency, I think acting on self-understanding is prior to this, as we shall see in chapter 5. If self-understanding is what is necessary for acting on reasons, then this is what is fundamental to agency. Briefly, this is because our self-understanding is often implicit, and emerges with our narrative understanding. This narrative understanding constitutes the perspective through which we understand our current context and through which we decide how to currently act. G&K's proposal goes beyond this to suggest that a minimal condition of agency is the ability to act in the future on plans that we currently make.

---

[20] For an indicator that our drive for consistency is socially learned and implicitly (and socially) practiced see McDowell on second-nature (1994) and Taylor on how we follow rules (1993).

[21] However see, for example Franklin (2015) and Dunn (1998) where different aspects of Velleman's theory are critically discussed

The idea that narrative self-understanding allows us to act for reasons prior to our ability to act on long-term plans is more consistent with some of Velleman's and Korsgaard's thinking than the line that G&K take. An account of agents as creatures that pursue plans and are consistent through time faces an objection. It appears, on this account, that people who undergo radical changes, which therefore interfere with pursuing such plans and goals, have their agency undermined. A response to this is given by Korsgaard (1989) who argues that while people may change, if the change is down to their own initiative, rather than imposed on them, this change itself is an act of agency. She writes, "you are not a different person *just* because you are very different. Authorial psychological connectedness is consistent with drastic changes, provided those changes are the result of actions by the person herself…"(p. 123). To fit this into my narrative account of agency, one could say that if it is some part of our own narrative understanding that drives reformulations in our narrative understanding, and hence our point of view, then there is a sense in which we are the author of the changes in our personality. I will give an example of this in chapter 5. Similarly, for Velleman, agency emerges with the desire to be consistent. So, if it is such a desire that is the driving force in a change in perspective, which may undercut diachronicity, then our agency is what defines this change rather than being challenged by it.

3. Prinz & the missing agent

3.1. Criteria for a response to Prinz

Before explaining G&K's objections to Prinz, let us recap briefly on what is at stake in this debate with Prinz. It is the particular line he draws between causation and constitution that is important for his argument. To counter his argument about emotions being sufficient for moral judgements, therefore, it is important that a reason is given for another process being constitutive of, and not just causally important for, moral judgement. However, the causal-constitution distinction is not something that G&K explicitly respond to. What I want to do after explaining

G&K's objection, is draw out why, read one way, it poses no problem for Prinz, but, if we dig deeper, we can unearth a set of commitments underlying their arguments that would pose a real problem for Prinz's framework.

In the last chapter we saw that Prinz thinks that emotions are the sole constituents of moral judgements but, nonetheless, there are important causal contributions from other psychological processes. By this he means that emotion is the only proper part of a moral judgement, but other processes play a role in producing that emotion. Specifically, deliberative processes are necessary causal components for moral judgements, but he doesn't seem to think that emotions are part of our deliberative processes, as we saw in the last chapter.

So, in the debate between Prinz and G&K, both sides agree that explicit rational processes are necessary for making moral judgements. Further, I want to highlight that Prinz's position allows that *any* processes, including MTT, could be necessary for moral judgements, while keeping his thesis that only emotions constitute emotions. Precisely because he utilises the causal-constitution distinction, he can respond that some process X is a necessary cause, but not a constituent, of moral judgements. G&K, in response are claiming that moral judgements are an exercise of our capacity for agency. If G&K are to have a substantial disagreement with Prinz, in addition to showing that,

i) Processes other than emotion are constitutive of moral agency;

They'd better also be showing that,

ii) There are substantial differences between their description of the processes that *constitute* agency and the description of various psychological processes *causing* an emotion.

Now we know what criteria we must use to assess G&K's response, lets look at the content of their response.

3.2. A problem for Prinz

It seems that G&K are making a similar claim about the neurosentimentalist (i.e. Prinz) and externalist account of moral judgement that Velleman makes about belief/desire descriptions of moral judgements. That is, neither the neurosentimentalist nor the externalist have accounted for our ability to act for

reasons. Moral judgments are activities of an agent. It is virtue of being an agent that one engages in decisions about what one ought to do, and to take part in this practice requires the desire to be consistent, and the ability to have reasons for acting. Both neurosentimentalism and externalism fail, G&K argue, because for reasons to exist for an agent a person must understand themselves as diachronic, and both tacit affective processes and rational rule application rely solely on synchronic processes. Diachronicity is needed, it is argued, because the ability to make decisions, and the ability to act on the decisions we have made, both depend on our capacity to understand ourselves as agents with pasts and futures.

In particular, G&K argue that a necessary component for being a moral agent is having the capacity for mental time travel (MTT). This includes episodic memory, which is the ability to subjectively experience past events (backwards time travel), and in turn seems necessary for us to imaginatively project ourselves into possible futures (forward time travel). So MTT is the ability to re-experience one's past and to be able to inhabit imagined futures. MTT is important, on their account, because our experience of existing through time enables decisions to mean something to us, so that we can act on them:

> *Agency requires the capacity for episodic memory (of events in subjective autobiography) and imaginative projection into the future in order for the subject to have the requisite inter-temporal perspective on her actions* (p. 588.)

In addition, such an agent has "detailed self-knowledge that is supported by autobiographical memory" (p. 601). A diachronic agent can act through time because they can experience their past and future *and* because they can reflect on who they are through using information in autobiographical memory.

G&K argue that both affective processing and procedural reasoning are by themselves insufficient for moral judgement, because, alone, do not enable us to have a personal perspective. A person with amnesia, who has only synchronic processes, cannot "represent the results of these processes to herself, because she is not a diachronic self but a bundle of habits linked to a synchronic reasoning system" (p. 596). We might wonder what this means. Here we refer back to Velleman. What it means for person to represent a decision to themselves, because

they are *not* just a bundle of habits, is that decision-making involves an agential perspective, which is the desire to be coherent and rational. It is through the desire to be coherent, along with an awareness of our selves that makes some actions coherent, that one gains reasons for acting. G&K's claim is that such a sense-making capacity is not present in either neurosentimentalism or externalism. So neither gives a place to the agent. While G&K fail to capture why this is the case, for reasons I outline shortly, I think this does turn out to be true of Prinz's neurosentimentalism. Externalism, as we saw in chapter 1, has been bracketed for the purpose of this thesis. So I will not be discussing their claims in regards to externalism.

G&K claim that neither neurosentimentalism nor externalism make room for the agent, since the mechanisms of moral judgements they propose are synchronic and therefore exclude the capacities necessary for moral agency. G&K diagnose the root of this problem to be that both neurosentimentalism and externalism rely on dual processing theories, but that this theory gives us inadequate tools for understanding moral judgements[22]. Both theories assume our cognitive capacities fall either in the 'tacit, associative, automatic and motivatory' camp (system 1 of dual systems theory), or the 'explicit, syntactical, effortful, and motivationally-inert' camp (system 2 of dual systems theory)[23]. Both theories assume that these processes cannot influence each other. The problem with both theories is that they only involve synchronic capacities, and so cannot make sense of how we can act for reasons[24]. That is, both theories ignore our capacity for

[22] We might question whether it really is dual processing theories that are the root of the problem, since sentimentalism and rationalist externalism predate these theories, and modern moral philosophers may not be that aware of contemporary cognitive science. G&K argue, however, that *neuro*sentimentalists explicitly rely on dual processing theories, and externalists implicitly do (p. 606). Perhaps, we might think it would be more accurate to claim that an older dichotomy between reason and intuition/emotion is a common cause of both dual processing theory and externalism.

[23] For examples of discussion and engagement with dual systems theory see Carruthers (2009) and Apperly & Butterfill (2009).

[24] We might wonder why diachronicity implies an experiential capacity. Maybe this is not a necessary connection. However, G&K think that it is this experiential quality that makes our diachronic reasons available for action. They talk of MTT as "an essentially indexical way of representing that information" (p. 601). I will expand on why we should see it this way in chapters 3, 5 & 6.

semantic autobiography and MTT.

G&K suggest that their,

> *Wider interpretation makes room for* diachronic *sentimentalists and rationalists who wish to incorporate a role for reflection and deliberation linked to a personal, intertemporal, perspective on action. The key is understanding the role played by the ventromedial prefrontal cortex in helping to create a moral agent constituted by the connection between deliberation, emotion, and diachronic self.* (p. 590, original emphasis.)

And they do not want to "dismiss the evidence about the role of tacit emotional and cognitive processes in moral judgement but rather to re-evaluate their role in constituting the moral agent" (p. 586). I pick up on this point here because I later want to suggest that G&K fail to give a determinate role to emotion.

So, similar to Velleman, G&K are claiming that we act for reasons when we act on our self-knowledge and diachronic self-experience. We can act for reasons when we understand, reflectively and experientially, who we are. It is through these processes that we can work out, and act on, what would make sense for us to do. Moral judgements are an activity of the agent, and agency is enabled by diachronic processes.

Because of the inadequacies of their theories, G&K argue that the neurosentimentalist and the externalist are forced to take implausible stances on whether to attribute agency to people who do not (or are impaired in their ability to) experience themselves as agents through time. For synchronic theories, people with this impairment qualify as making moral judgements. Such cases include people with amnesia who lack the autobiographical database (i.e. awareness and knowledge of their past), and people with ventromedial prefrontal cortex (vmPFC) damage who cannot make use of it.

A case study of a patient, called M.L. – who had memories but didn't feel a personal sense of connection to them – described him as irresponsible and a poor decision-maker. M.L. knew about his past but he could no longer mentally time travel, he didn't re-experience his past. It was observed that M.L. had difficulties acting as a responsible parent for his children, and that he needed supervision. Another vmPFC patient, called EVR, took a long time to make decisions, for

example what restaurant to eat in, and reported that even when he made a decision, he found it hard knowing what they meant for action[25]. G&K note that the characteristics associated with vmPFC patients include impulsivity, inability to plan, and a lack of motivation to act on any decisions they do make. It seems to G&K that we should not deem people who are impaired in this way moral agents because they lack reasons for acting that guide their future actions. Similarly, if someone has amnesia and has no ability to locate themselves in a personal history, they both appear to lack the information needed to make a decision, and also to carry out those plans.

Both people with amnesia and people with vmPFC damage seem to have impaired agency insofar as they have problems deciding what to do, and acting on their decisions:

> *Patients with ventromedial damage and amnesia either fail to make [moral] judgements or are impaired* as agents *in their inability to choose and act in accordance with reasons they accept… But this rationalist position cannot be mapped onto a dual process model which sees action as either driven from below by automatic affective processes or from above by procedural reasoning detached from affect and personal motivation…*(p. 596, original emphasis.)

So, for a synchronic theorist to account for the data they seem committed to saying either that people with these problems *are* agents, despite their problems, or that they make moral judgements, but are not agents[26]. On the other hand, a diachronic

[25] Craver (2014) argues that people who lack MTT do not have impairments in their diachronic understanding. Further, there is evidence that one person, K.C., who lacks episodic memory but not semantic memory, retains good decision-making abilities (Kwan et al., 2012). This may seem like a problem to G&K. For they tie our capacity to MTT to our ability to make decisions. However, I think there are reasons to be suspicious of what K.C.'s results mean for agency. The experiments showing this involved a gambling paradigm, which involves a relatively simple cost-benefit analysis, compared to the contextually rich, ethically nuanced, everyday decision-making. That is, it is not clear that these experiments are ecologically valid. Additionally, these experiments ask participants to make a choice among hypothetical scenarios, so they are not required to act on their decision in the distal future. So, it is not clear how these experiments relate to agency as G&K are understanding it here, where agency involves the ability to act on judgements we have made in the past.

[26] There may be reasons other than those that G&K suggest that mean we do not see these people as agents. However, as we shall see shortly, I think there alternative reasons for arguing that neurosentimentalism fails to give an account of agency.

theory can account for this phenomenon by claiming that these people cannot make moral judgements because their agency is impaired.

These cases indicate to G&K that our capacity to mentally time travel in particular is necessary for being an agent, for being a creature that can act responsibly. They say that, "it is not just loss of information about future consequences but of an essentially indexical way of representing that information which impairs deliberation" (p. 601). MTT enables us to experience ourselves as existing as diachronic creatures, and therefore to experience our projects, commitments and values as ours.

But synchronic theories, it seems, cannot account for this data because they don't take into account that moral judgements require the diachronic capacities that enable agency. So the synchronic theorists could stipulate that people with these problems have no impairment in agency, which doesn't seem true, given the problems people with vmPFC damage or amnesia have with acting responsibly. Alternatively, synchronic theorists can claim that people with vmPFC lesions or amnesia can make moral judgements, but that such judgements do not require agency. There are two problems with this. First, at least some people with vmPFC lesions seem to be impaired in making judgements, and second it is unclear what a moral judgement is if it is not an exercise of agency, i.e. if it is not something that we can act on.

So, this is meant to be a problem for Prinz. Presented from the angle that G&K stress, it is related to the dual systems account of cognition. The dual systems account, which splits cognition into immediate affective and intuitive processes on one side, and explicit deductive reasoning on the other, leaves out a place for experiential memory and our self-knowledge gained through autiobiographical memory. On this story, once we reject dual systems, and give a place for diachronic capacities, we rehabilitate the agent, and we reject neurosentimentalism. Emotion (alone) cannot be constitutive of moral judgements, because moral judgements are an activity of an agent, and an agent is constituted by their diachronic and deliberative capacities. We can see this when we look at the behaviour of people who lack the experience, and knowledge, of themselves as creatures that exist through time. But this isn't going to work or, as we shall see, it will only work with

a few more ideas made explicit. The real problem for Prinz is something else that G&K have expressed but not explained, and that concerns the confusion between two modes of explanation. Focusing on the dual systems account obscures rather than facilitates our grasp of what is wrong with Prinz's account.

3.3. Not actually a problem for Prinz?

To see that Prinz can accommodate G&K's objection we need to remember a signature move: making a distinction between a cause of a moral judgement and a constituent of it. Importantly, Prinz has already conceded that deliberative reason is a necessary cause of moral judgements. It is in this sense that he thinks that our moral judgements are merited. We need to take a look at how and why he deals with this issue to 1) see how he can deal with G&K's objection and 2) show how there is room to suggest that he could be confused in this style of thinking.

Prinz believes that moral judgements are merited, not merely caused, because we have a type of control over our emotions that we do not have over all psychological processes, such as that of seeing a colour. Prinz (2007) asks,

> *Where does this intuition of meriting come from? It may have something to do with the fact that the link between emotions and their causes is less direct than the link between colour and their causes. In the colour case, the causal connection is fixed by our biology…In the case of emotions, we can exercise more control…*(p.114.)

It is because we have control of our emotions, that, when it comes to moral judgements, we think someone can be held accountable for getting them wrong:

> *We cannot hold a colour-blind person accountable for failing to distinguish red from green, but we can hold an emotionally healthy person accountable for being afraid of foreigners.* (ibid.)

So "moral emotions are not merely caused; they are merited by their causes" (p.115). Prinz's criteria for a psychological response being merited concerns whether "a) it applies to its cause, and b)… the agent can be held responsible to some degree for failing to have that response" (p. 114).

Despite all this, Prinz continues to argue that emotions are the only constituent of moral judgements. And he does so by arguing that rational deliberation is an important cause of a moral judgement, but not a constituent of it.

As we saw in the last chapter, his reasons for doing so appear to be mainly motivated by a commitment to sentimentalism, combined with evidence from dumbfounding experiments. In dumbfounding experiments, deliberation doesn't alter one's moral judgements, and emotions and moral judgements appear to co-vary. Therefore deliberation doesn't appear to be a constituent of moral judgements, but emotions do.

Similarly, Prinz can respond that MTT may be an important cause of the emotions that constitute moral judgements but that we needn't see MTT as a constituent. Just like deliberative reasoning can feed into our judgements, so can other psychological processes. This way of putting it may appear to fit with the role G&K give to MTT. For example, they claim that the problem with amnesia is that it impairs someone's ability to "assemble and inhabit the episodes necessary to support adequate deliberations" and so, "impairs [a person's] capacity for agency" (p. 589). One could read this as MTT being an input to our deliberative processes, it is food for thought. Put this way, there seems to be some convergence between G&K's account and Prinz's, where one can understand moral judgements as a result of a causal chain constituted by deliberative reasoning. And, whatever mental process one puts forward as being important for moral agency, it seems that Prinz can make the same move. Because it remains true that if we see a judgement as the result of a merely causal process, then it is possible that a single emotion constitutes that judgement, while other psychological processes are crucial causal inputs that help shape that emotion.

Prinz takes it that he has shown that emotion constitutes moral judgement, and that all deliberative rational processes are causes. What we are looking for, in G&K's account, is a reason why it is not the case that deliberative processes are only causally important (as suggested in criterion ii.). However, they don't appear to provide us with this. What they do give us is a reason to think that certain diachronic processes are necessary for moral judgements. But, like the sun is a necessary cause of sunburn, but not a constituent of it, showing that a process is necessary for a mental state is not the same as showing that it is a constituent of that mental state. If G&K want to go further, and claim that diachronic processes constitute moral judgements, they have to say something more.

And, without explaining why, it is not enough for G&K to say that because diachronic processes are constitutive of agency, and it is an agent that makes a judgement, therefore diachronic processes are constitutive of moral judgements. Remember that what G&K are focused on is how to characterise agency, and they understand judgements as activities of an agent. But Prinz has argued that agents are rational creatures, in the sense that they can deliberate and can justify, or fail to justify, their beliefs. Nonetheless, he thinks this is consistent with saying that the processes that make us an agent are only important causes of our moral judgements, not constituents of them. So he could make an analogous response to the addition of diachronic processes: sure, they constitute us as agents, but our agency is a cause of our moral judgements[27].

Furthermore, if emotions are merited because they are under an agent's rational control, then Prinz can reply that he hasn't made the mistake that G&K accuse him of. On his account, it looks like we do act for the reasons, in the sense that emotions represent values that are under rational control. Now he can add that he missed a crucial component of what matters for an emotion to be under rational control, and that is that our emotions are also shaped by MTT. Our reasons for acting are constituted through emotions, but are shaped by our self-knowledge and MTT, and controlled by our deliberative reasoning. If agency is constituted by our self-understanding – born through deliberation, our self-concept and experience of our self through time – then, on one reading, Prinz can reinstate the agent, but remain, fundamentally, a sentimentalist. Many processes are required for a moral judgement, but only in the sense that they cause an emotion, which is what constitutes the moral judgement.

We can understand Prinz's concern here in light of G&K's commentary on the moral judgement debate. Part of what drives Prinz to note that emotions are merited is some concern for understanding moral judgements as an activity of an

[27] Of course, G&K could accept a state of stalemate, and argue that none of Prinz's reasons for seeing emotions as sufficient for constituting moral judgements are decisive. Therefore, since they find it more compelling to see diachronic capacities as constitutive of moral judgements, they are going to continue in that vein. However, I find this unsatisfying. Such a position fails to articulate what commitments one would have for making a process constitutive rather than causal.

agent. We can see this, because in his discussion, Prinz refers to the need for his theory to account for agency, and this is why he argues that emotions are under rational control. By describing emotions as under rational control, Prinz provides a way to understand moral judgements as the result of agency. It is our rational control that allows a judgement to be seen as ours, rather than something that happens to us. And it is for this reason that we can understand our moral judgements as something we are responsible for, as Prinz does: we are responsible for them because they are an act of agency.

Prinz does not *have* to accept that diachronic capacities are important for a judgement being merited. It is open to him to show that G&K (and Velleman & co.) are mistaken in this. However, on first glance, it appears that Prinz does not need to go as far as rejecting G&K's insights for him to defend his sentimentalist theory. He can accept that MTT and autobiographical memory contribute to our judgements being merited while still denying they are constitutive of moral judgements.

But put this way, it doesn't look like G&K have met criterion ii) for responding to Prinz. We have yet to see a good reason for not recasting their claim in terms of a causal process that leads to an emotion, where that emotion is the only constituent of a moral judgement.

3.4. Prinz and losing the agent (again)

What I want to do now is explain what I think needs to be made explicit for criterion ii) to be met. This involves explaining a potential confusion that Prinz has made in his reasoning, and explaining how this leads to a dilemma where he either loses the agent, or has to reconsider his sentimentalism. For this confusion to become evident, however, requires explaining a competing conceptual framework to the one Prinz uses, which has been proposed by McDowell.

The first thing to note is that it looks like our intentional and epistemic behaviour is normative, in a way that our causal explanations are not. It is hard to see how one atom affecting another, or one neuron stimulating another, can result in something being justified or unjustified. No matter how complicated the causal

chain is, it still looks like we've explained how things happen, rather than what is justified.

Justification has normative force. It tells us what we ought to do or believe given some state of affairs. But it is hard to understand the relevance of a description of mechanism to questions of 'ought'. It appears that there is no clear way to get from causal explanations to epistemic explanations.

While Prinz wants to retain emotions as epistemic processes, by arguing they are conceptual, we saw in the last chapter that he does not think they are participants in our deliberative processes. Emotions can only be caused by deliberation. Because Prinz implies that emotions are not participants in deliberation but caused by it, he implies that we should understand emotions in causal terms. In chapter 1 we saw that, on his own definition of concepts, Prinz has a tension where he both sees emotions as conceptual and not conceptual. Moreover, there is another switch, on Prinz's account, from the space of reasons, to causal explanations. We reason, and that triggers an emotion. All the while, he maintains that the caused state is a judgement. What I want to argue now is that if we apply McDowell's framework to Prinz's theory that moral emotions are concepts, but not part of deliberative reasoning, new issues arise.

On McDowell's framework, we are led into confusion if we understand an emotional response to reasoning as simultaneously a judgement and a reaction to reasoning, rather than itself part of our abilities to justify, infer and clarify. We are led into confusion because our agency, and with it our ability to judge, is characterised by our capacity to engage in what McDowell calls the "space of reasons". The space of reasons, in turn, is characterised by relationships of justification and their lack, not by causal and mechanical processes (see, for example, Lecture IV, 1994). According to McDowell, then, if emotions are judgements, they should be understood as part of our deliberative capacities.

However, because emotions are caused, but not participants in reasoning, on Prinz's account, when he understands them as judgements, it looks like he has mixed together two types of explanation. For McDowell, when we mix together talk of merely causal processes with talk of intentional and epistemic behaviour, we

have mixed together two distinct modes of understanding, and in doing so, have applied one type of rule to domain where that rule doesn't apply.

This observation is similar to Sellars' (1956) argument against the way that sense-datum psychology is used in epistemology. To explain "sensory perception in scientific style" (p. 132-3) and simultaneously hold that it has an epistemic character, endowing us with awareness of how the world is, creates a "mongrel resulting from the crossbreeding of two ideas" (p. 132). Emotions, as endowing us with an awareness of moral properties, on Prinz's picture, are simultaneously held to be mere causal processes and have an epistemic character. That is they are, on the one hand, epistemic and intentional, and other hand, merely causal.

This is not to say that causal processes do not *enable* us to think inferentially, just that when we understand some creature as an agent, we are taking a particular explanatory stance towards them, one where we understand them as having the capacity to engage in the practice of logical inference and justification. This is another distinction that McDowell (1994) makes, where a *constitutive condition* is the best way to characterise what makes a thing a thing of a particular sort, and an *enabling condition* is the mechanism by which a thing of that sort can exist. So what *constitutes* agency is being a creature capable of standing back and asking things such as, 'given I believe x, am I justified in believing y?', or 'given the way the world is/I am, what action is justified?'[28]. But what *enables* this is a particular physical substrate, which involves (in the human case) neural machinery, and perhaps the non-neural body and the world.

To see the claim that we should understand an agent as operating in a space of reasons as legitimate, we must abandon what McDowell calls "bald naturalism". Bald naturalism is the view that for something to be explained naturalistically, it has to be explained through the laws, and patterns of cause and effect that the natural sciences study and explain. If we take bald naturalism to be the only mode

[28] Note that the argument is that what is crucial is our *capacity* to stand back and reflect, not that we always engage this capacity before acting. I will return to this in chapter 5. Briefly, however, I mean that it may be that we often act without reflecting, for example I cross the road when the green person lights up without thinking, 'the green person is a sign I can cross the road, they have lit up, so I can cross the road'. However, if you asked me why I crossed the road when I did, my reasons would be available to me.

of explanation that enables us to naturalise a subject, then talking of teleological creatures, for which there are a set of inferentially justified reasons, should simply reduce to causal explanations without any loss of understanding of the subject we are trying to explain. Bald naturalists[29] think that we will be able to understand normative terms like 'justify' through purely causal mechanisms.

However, those that want to break with bald naturalism, like McDowell, have reservations about these claims. First, we don't have a story about how causal processes explain normative practices yet, and it is hard to see how we ever will do. As mentioned above, there just seems to be a conceptual leap when going from a causal explanation to normative explanations.

Additionally, there doesn't seem to be a need to reduce explanations that use the space of reasons to causal explanations unless we are bald naturalists. And it isn't evident that we need to be bald naturalists to be naturalists.

McDowell defends the ability to be a non-bald naturalist by appealing to the practices and patterns that govern how we live in daily life (1994, p. 78-79). These practices are practices of justifying our behaviour and beliefs, and asking others to justify theirs. And this behaviour, of giving and asking for reasons, is characteristic of the kind of animals we are, for McDowell. Because it is a type of animal behaviour, it is a part of the natural world:

> *On this view, our lives are animal lives through and through; it is just that we are animals of a rather special kind. The capacity to engage in rational reflection about how we should live belongs to our nature, as the kind of animals we are, no less than, say, the capacity to walk on two legs.* (2008, p. 215.)

Explanations that refer to this behaviour, therefore, are naturalistic explanations:

> *By dint of exploiting, in an utterly intuitive way, ideas like that of the patterns characteristic of the life of animals of a certain kind, we can insist that such phenomena, even though they are beyond the reach of natural-scientific explanation, are perfectly real, without thereby relegating them to the sphere of the occult or the supernatural.* (ibid., p. 217.)

If explaining the patterns of human behaviour is a naturalistic explanation, and we think that understanding humans to engage in 'a space of reasons' is a good way of

[29] Not to be confused with bald naturists.

characterising such behaviour, then explanation that refer to the space of reasons count as a naturalistic, rather than supernatural, explanations.

McDowell also has a general idea of how it is that we become able to reason. We, as social animals, come to be capable of reasoning, for McDowell, because of the cultural practices we are taught as we grow up. So, for McDowell, the ability to take part in the space of reasons is part of our 'second nature'. That is, it is a part of the natural world that depends on our cultural practices.

The outcome, for McDowell, is that we can have a type of naturalism that encompasses the space of reasons. But note that this comes with the claim that the space of reasons is of its own kind (*sui generis*), it is governed by normative relations, which are not found in the causal mode of explanation. The type of naturalism that McDowell is proposing includes the type of explanations where one can be right or wrong, justified or unjustified, and where we take creatures to be free and not merely caused[30].

So, another reason why we might adopt McDowell's naturalism, as opposed to bald naturalism is that we no longer need to have a story about how our mental activity is *constituted* through causal processes. It's true that we no longer have this story, but that's because we have no need for such a story. Because on McDowell's account we can still give a causal story about what *enables* the space of reasons to exist. That this enabling story cannot explain how causal processes could have a normative character is no longer a problem because it is assumed that causal processes enable, but are not characterised by, normative relations.

What I want to remind us of now is that Prinz has some commitment to explaining how moral judgements could exist in the space of reasons. This is what is implied when he expresses the view that the emotions that constitute moral judgements can be merited, and his commitment to seeing moral emotions as constituting moral concepts. 'Merit' acts like 'justification' in this argument, in the sense that Prinz is arguing that emotions can constitute judgements in the sense

[30] This ought not imply being a libertarian. The idea is that there is a causal-mechanical *enabling* explanation of how it is that we are free, but that when explaining what *constitutes* us, we use a different type of explanation. Because what constitutes us as agents are our intentional, and epistemic, practices.

that we *ought* to have those emotions, because they are caused by our deliberative processes.

It is also Prinz's concern with making emotions a result of agency that underlies his concern with control: emotions are under our rational control, and this is what allows us to understand moral judgements as a result of our agency, as something we do rather than something that happens to us. Such an activity is what makes moral judgements the result of a creature that is free (or what McDowell would call "spontaneous"), and therefore responsible for its actions. And to be clear, Prinz can hold that we act for reasons while also holding that some causal psychological process enables us to participate in the space of reasoning. In this way, it is consistent for Prinz to claim that moral judgements are merited as well as caused. Moral judgements are merited, by reasons, and they are enabled by some causal process.

According to McDowell's view, it is a confused line of reasoning to suggest that the right type of cause can merit some judgement X. The right type of justification can merit some judgement X, but a process understood in purely causal language cannot. Our reasoning can merit an emotion if, and only if, the relation between the reason and the emotion is one of justification. Only if we *ought* to have that emotion given that reason. But this then brings emotion, or at least those that are justified by reason, into the space of reasons. Emotions, on this account, would (have the potential to) participate in deliberation, and not just be elicited by it. Because, for McDowell, it is conceptual processes that participate in reasoning, emotions would have to be unambiguously conceptual to participate in reasoning. This presents a problem for Prinz. For he does not appear to think that emotions participate in reasoning.

It might be suggested that I am no longer paying heed to Prinz's definition of what it is to for an emotion to be merited. For him, an emotion is merited if it is caused by good reasoning, not if it can participate in deliberation. However, as argued above, there is a problem with understanding 'merited' this way. If we look at what Prinz means by 'judgement', it is the type of mongrel understanding where moral emotions are simultaneously understood to play the epistemic role of representing what is good (for us) and to be triggered, rather than part of,

deliberative processes. One way to put this is that Prinz wants to hold that we should understand what *constitutes* emotions as simultaneously a type of causal reaction and an instance of moral knowledge. What is being contested above is whether this is clearly the case. As has been suggested, it is not yet clear how moral knowledge can be characterised as a type of causal process, and therefore it is not *yet* clear how moral knowledge is constituted through a causal process. Further, it looks like there might be a schism between causal explanations and epistemic explanations, making it unclear how we could *ever* hold this. Finally, if we drop bald naturalism and use McDowell's framework we have no need to find a way to hold that we could characterise moral knowledge this way.

This brings us back to why Prinz cannot simply adopt the position that emotion and deliberative reason independently, but jointly, constitute moral judgements. If we cannot characterise emotions as conceptual, in the sense that they are capable of participating in deliberation, then, on McDowell's account, they cannot be understood as participating in the activity of judging. On this solution to Prinz's problem, it now looks like only the rational deliberative part of a judgement is constitutive of judging. This requires Prinz to abandon sentimentalism, which is a reversal of his position.

Because emotions, for Prinz, motivate action, this would also require that he abandon his moral internalism. This presents a different problem for our characterisation of a process as a judgement: how do we understand something as a moral judgement if it not something that an agent can act on?

A question arises here about how moral internalism can escape mixing causal explanations with epistemic ones, given that its claims that judgements motivate actions seems to imply that judgements can cause actions. My internalism, however, is a commitment to the idea that our characterisation of moral judgements and agency ought to allow for an enabling explanation of how agents can act on reasons.

If emotions are judgements in McDowell's sense, then we have to understand them as the type of things that can participate in being justified by a reason and justifying an action, rather than being merely caused and be a cause. For both Prinz

and McDowell, an activity that can be justified and justifying is conceptual. So, if we are to say that emotions are justified (or 'merited' as Prinz puts it) and justifying, then, according to McDowell, emotions must be conceptual. In some ways, this is the case for Prinz too, in that it is concepts, for Prinz, that participate in rational activity, rather than being merely caused by it.

However, if we adopt McDowell's reasoning, then Prinz's description of events is muddled – he is moving between the space of reasons and a causal description. He is marrying two inconsistent types of description, and applying the concept of 'merit', which is an epistemic and normative concept, where it does not apply.

Within this way of understanding 'judgement' Prinz is presented with a dilemma. If he wants to maintain that moral emotions constitute moral judgements, he can either change his basic understanding of moral emotions, so that they become fully conceptual (i.e. involved in deliberation), and keep the agent, or he can keep his basic understanding of moral emotions, and lose the agent.

The first move is the one I am going to make. In chapter 5, I will argue that emotions do constitute concepts, and explain how it is possible to hold this. Prinz can also make this move if he wishes, but it is not a claim that can be made without further substantiation, because it flies in the face of an everyday assumption. Namely that emotions and rationality are opposed to each other.

The second move would mean that G&K's argument continues to have a hold on him. If Prinz is committed to bald naturalism, then it is no longer clear that we should take moral emotions as moral judgements any more, because they cannot be part of deliberation, so they cannot be the types of things that are justified. They can be causes of behaviour, but they no longer look like the right type of thing to understand how it is that we can *act*.

On McDowell's framework, as long as we are committed to explaining how it is that we can act for reasons, something more like the first move is required. That is, we have to understand emotions as being in the space of reasons.

Seen this way, Prinz's switch between modes of explanation means that his causal-constitution argument is either opaque or wrong. Both possibilities

undermine his argument that emotions, as opposed to our deliberative capacities, are constituents of moral judgements. First, it doesn't seem we can make sense of his argument any more. The space of reasons is characterised by relations of justification, not causes. Second, if we are more careful about how we use modes of explanation, then it looks like his claim that emotions, rather than deliberation, constitute judgements is false: either our emotional capacities are conceptual, in the sense that they can participate in deliberation, or we have no understanding of how we make moral judgements at all.

We can recharacterise Velleman's account using this framework too. Self-understanding should not be characterised as a particular part of the causal order that allows us to act for reasons. Instead, there is some causal process that enables self-understanding. Self-understanding is part of what constitutes the space of reasons. It provides a reason for acting because, since we want to be consistent, it justifies some courses of action over others.

In chapter 5 I will present my solution to the dilemma Prinz faces. I will argue that moral concepts, which participate in deliberation, are constituted through emotions, and that it is the interrelation of moral concepts that acts as a justification for action. Moral concepts justify only as part of broad narrative networks that constitute our values. In this way, I situate the argument above within a more detailed theoretic framework of how we make moral judgements, and include phenomenological and empirical support for this picture. In doing so I will commit, unlike Prinz, to the idea that moral concepts are part of deliberation, and depart from sentimentalism. Or, maybe more precisely, my theory is one where rationalism and sentimentalism converge.

In chapter 5, I will also respond to potential criticisms of the argument I have presented here: that it is hyper-rational because it means that we only act when we have reflected prior to the action; and that it gives us too much power, because it implies that it would be simple to articulate, understand and change our reasons. Briefly, I don't think either of these characteristics are necessarily implied by the picture I have presented here.

4. Aims and methods

4.1. Naturalising agency

So we have seen how, although G&K's response to neurosentimentalism does not necessarily pose a problem for Prinz, their emphasis on agency can lead us in a direction where a response is provided. I have argued that ideas of agency can be understood as belonging to a type of explanation that is not congruent with Prinz's explanation of moral judgements.

However, another question then arises of what contribution G&K take themselves to be making to a positive account of agency. In this, I think they share similar commitments and aims to Prinz, ones that are adopted in this thesis. One of these shared commitments is a sensitivity to empirical literature, as one of the considerations in play when deciding what theory of moral judgements makes most sense. Before returning to an outline of my general methodology, I want to explain a couple of projects I will be adopting from G&K.

The first is G&K's concern with providing psychological and neurological processes that they take to enable the type of self-understanding required for agency. This is a task that I will be taking up for the rest of the thesis. We can see evidence of this approach, when, in their explanation of acting for reasons, they state that,

> *In so committing ourselves, as Velleman (1991 and 1997) points out, we provide reasons for ourselves in the future, reasons which will be ours, but which we would otherwise not have had. In this way we construct ourselves as particular, temporally extended, agents. Our diachronic reasons, made salient to us via our capacity for mental time travel, are thus in a position to compete with synchronically occurring wants. In effect they become normative for us.* (2010, p. 602.)

Part of G&K's project is to use the psychologically informed theory of MTT, and the neurological evidence about the circuitry involved in it, to explain what enables us to have self-understanding, and therefore have reasons for acting. Their argument is that MTT and autobiographical memory are necessary for us to have the self-understanding required to create and commit to projects across time. And it is for this reason that people with amnesia and ventromedial damage are

impaired in this activity. The (neuro)psychological literature on MTT then allows us to provide the enabling conditions for Velleman's theory of agency.

Part of what I will do in the proceeding chapters is challenge the way that G&K interpret this evidence. I will claim that the vmPFC should be understood as enabling the more general capacity for narrative understanding, where MTT is a particular type of narrative understanding. It is through narrative understanding that we come to understand our selves and the world, what we value, and through which we gain reasons for actions.

A large component of this endeavour to provide an account of what psychological processes make self-understanding possible will be to explain why we should see emotion as integral to this process of narrative understanding. Finding a role for emotion is something that G&K have an explicit commitment to, but it is unclear what role they give it in their account.

While they mention that memory and imagination "involve the activation of relevant perceptual, sensory and *emotional* systems in the absence of environmental stimulus" (p. 599, emphasis added), they do not expand on what role emotion has here[31]. They also suggest that part of the problem that ventromedial patients have in making decisions is that they "make impersonal judgements without experiencing the conflict produced in normal subjects by personal and emotional aspects of a situation" (p. 588). Again, they do not expand on why emotions are useful for decision-making. They do, however, mention that a lack of emotion is responsible for ventromedial patients finding it hard to form judgements that motivate them to act (p. 605). In their conclusion they also mention that a diachronic sentimentalist may argue that emotion is "recruited when we imagine another's situation". But this could well be understood as emotion having a subsidiary role in moral agency, which at heart is constituted by

[31] However, they do expand on this further in their 2017 paper. Here they make explicit what they understand the role of emotions in MTT to be. Similar to me, G&K argue that emotion is essential for the subjective experience of MTT, the sense that it is inhabited. This is central to its motivating character. However, unlike me, they do not expand on this to provide an elaborate theory of moral cognition and agency that centers on emotion.

our explicit memories and imaginings. So their commitment in including emotion as a constituent of moral agency appears unfulfilled.

What I will do in the proceeding chapters is explain how emotions are crucial for the embodied perspective characteristic of narrative understanding, and therefore MTT. And it is the embodied nature of narrative understanding that enables our understanding to be intrinsically connected to our ability to act.

4.2. General methodology

We've now gathered together the main commitments that I will be inheriting from Prinz and G&K. Both are committed to moral internalism, giving some place to emotion in agency, and using empirical evidence as one source to assess a theory of agency by. From G&K, I will be continuing their endeavour to give an account of what processes enable the self-understanding through which it is possible to act for reasons. However, how we understand 'self-understanding' will go through a transformation. In particular, we shall see that self-understanding will no longer necessarily imply the activity of thinking about our own psychological processes.

The project I am about to embark on, is arguing that there are a multitude of reasons that at once speak against Prinz's account of moral judgements and in favour of an account of agency centred on self-understanding. The methodology used to argue this will rarely involve straightforward refutation of Prinz, nor an obvious winning argument in favour of an alternative account. Instead what will be presented is a network of considerations, empirical and conceptual, and the proposal is that Prinz's theory cannot make sense of these considerations. But we can best account for these considerations if we adopt the theory that embodied narrative understanding is constitutive of agency. Such an account adopts an alternative set of conceptual considerations from Prinz, namely the considerations that McDowell presents us with above. Because of this, the alternative to Prinz I will be presenting does not meet him on his own terms, and therefore shifts the debate rather than directly undermining his position.

We may wonder why I should switch grounds, whether I could translate my account into a brute naturalist story, whether I can remain on Prinz's ground while simultaneously rejecting his type of sentimentalism. While such a translation may

be possible, that remains to be seen. Further, McDowell's observations bring into view the possibility that Prinz's ground is not one where an explanation of agency or judgements can be provided, because it could render incomprehensible how my proposal relates to agency and the activity of judging.

My alternative to Prinz integrates important insights from various thinkers and shows how they are consistent, provides various novel insights into how we can understand agency, and makes sense of a wide range of empirical findings. So one major drive in the argument is explanatory force: if we understand these things this way, we can explain and make coherent various lines of thinking.

One strategy used will involve presenting it as a virtue if a theory can make coherent relevant phenomenological observations and particular empirical findings. However, why this should be a reason for theory choice is not obvious. It is not obvious at all that phenomenologists should care about what the science says, nor that scientists should care about the phenomenology. So it worth explaining why we should think that their consistency matters.

One explanation comes from Wheeler (2013). Wheeler argues that, while neither discipline may decisively prove the other right or wrong, they mutually inform one another. We can understand how this is possible, he suggests, if we apply McDowell's distinction between constitutive understanding and enabling understanding. As we saw above, constitutive understanding is reached through the activity of describing and refining the conditions that make some phenomenon the kind of thing it is. Enabling understanding is reached when we can give some causal story for how that phenomenon could be generated. Phenomenology can contribute to our constitutive understanding of experience, neuroscience can give us enabling understanding of how experience is possible. Specifically, phenomenology can give us constitutive understanding of the transcendental kind. It does not just give us information we gain from everyday introspection, but information gained from careful analysis concerning the conditions that make experience possible.

The science, then, takes its cue about what it is looking for partly from phenomenology. Phenomenology says: here's a description of the conditions that

make experience possible, and neuroscience responds: ok then, I'll see if there's a neural mechanism that would explain how those conditions could obtain.

In return, the phenomenologist is constrained by science in the sense that phenomenological descriptions shouldn't be mysterious. That is, there should be some scientific explanation possible for how it could be that the descriptions phenomenology gives of experience could be realised by stuff that science can explain. If our best science cannot offer an enabling explanation, then phenomenologists should at least reconsider their descriptions of the transcendental conditions for experience.

Further, scientific evidence may provide positive evidence that allows phenomenologists to reconceive what the best descriptions of the nature of experience is. Wheeler uses the example of research in situated robotics that apparently showed that one can make nuanced distinctions between types of representations such that some prereflective experience can be understood as involving minimal representations that help direct attention, without needing rich representations. Hence phenomenology and science are in a relationship of semi-dependence, where they can mutually influence one another, and yet neither holds absolute authority over the other.

Consistency between these two disciplines, then, is taken to be a good sign because their agreement gives us some reason to accept both the phenomenology and the science. That is, when in agreement, they lend each other mutual support. We know that our explication of the conditions of experience can be made sense of through science, and we know our science can be made sense of through our phenomenological explications. We have a more complete and rounded sense of the world when these two things are in harmony.

This is a way of expressing the more general methodology too. It is not just the fit between phenomenology and empirical evidence that will be presented as a virtue, but the fit between other logical or conceptual considerations – which serve to further our constitutive understanding of agency – and the empirical evidence.

This form of argumentation makes the structure of this thesis complicated. The argument takes a web-like form, where issues interconnect in various ways,

and are therefore revisited, but for different purposes. This is another way of saying: I will be grateful for your patience.

5. Conclusion

This chapter has surveyed the response of G&K to Prinz's sentimentalism. I have used it to pull out the underlying criticism of Prinz's argument that deliberative process can be important for causing moral judgements but that they do not constitute such judgements. First, it looks like their argument poses no real challenge to this claim, but serves to embellish what causal processes are necessary for a moral emotion to be produced. However, I have argued that once we think more carefully about agency, we find a more fundamental criticism of Prinz's set-up, where he appears to moves seamlessly between two inconsistent modes of understanding. On this understanding of agency, made explicit by McDowell, Prinz's claim that emotions are causally related to deliberation and yet constitute moral judgements becomes opaque. If he makes his reasoning consistent according to McDowell's framework, he is left in a situation where he either has to accept moral emotions are fully conceptual, or that he has not given us an account of moral judgements.

Additionally, I have used this chapter to introduce Velleman's account of agency as requiring self-understanding, a claim that G&K see themselves as providing a psychological and neurological account of. While I will adopt and adapt Velleman's account, and I inherit G&K's methodology of finding an enabling story for it, in the next chapter I will start to question the role that they give MTT and the vmPFC. Furthermore, I will begin the enterprise of explaining why emotion must be a part of the explanation of why people with vmPFC have the issues with decision-making that they have.

Building an alternative theory of moral judgements to Prinz will take some time and patience. So, in the next two chapters, I bracket Prinz, and instead begin developing my alternative. In the next chapter I respond to G&K's claim that mental time travel is important for making and acting on responsible decisions. In

chapter 4, I put this back into the context of agency. I develop this into an account of moral agency in chapter 5, where Prinz returns to the scene.

Emotion in Narrative Understanding and Mental Time Travel

Are all of us the same, I wonder, navigating our lives by interpreting the silences between words spoken, analysing the returning echoes of our memory in order to chart the terrain, in order to make sense of the world around us.
Tan Twang Eng, The Garden of Evening Mists, 2012, p. 323.

1. Introduction

Mental time travel, as we saw, is the ability to imagine oneself in the future and re-experience the past. It evokes sensory experiences: we may visualise the mischievous smile on a child's face, imagine the jarring tone in our teacher's voice, or feel the rhythm of the dance we waltzed last night. Mental time travel (MTT) is also experiential in the sense of being emotional: memories of that mischievous smile warm our heart, of that imagined tone of voice come with being annoyed and anxious, and remembering that dance is partnered with elation. This observation forms the basis of much of the rest of this thesis. For this chapter and the next two, we will be concentrating on the affective dimension of narrative thinking, in chapter 6 we will return to see why this cannot be separated off from the sensory aspect.

Gerrans & Kennett (G&K) claim the capacity for MTT is necessary for being a moral agent. This is because MTT provides us with the ability to subjectively inhabit our past and to imagine our future. In turn, this is necessary for an agent to have reasons for acting that are normatively felt because "engaging in this normative domain just is the process of learning to transcend the present moment, both cognitively and behaviourally" (G&K, 2010, p. 602.). To support this idea G&K make reference to the philosophers Velleman, Bratman & Korsgaard, who express the view that to be an agent is to have reasons to act, and one has reasons to act when one tries to act ways that are consistent with one's values. MTT, for G&K, is that capacity that allows one to have an autobiographical context to one's decision-making. We use this context to make decisions about what we have reasons to do, because it is with this information that we form our ideas of what matters to us, and how we want to lead our life.

Furthermore, G&K are internalists about moral judgements: they think moral judgements necessarily motivate us to act. 'Reasons for an agent' must be reasons that have some motivational force. For them MTT provides us with a means to have 'reasons for actions' not only in the sense that it helps us make decisions about what to do, but also because it is involved in motivating us to do what we have decided to do.

What is crucial for G&K is that MTT provides us with more than semantic knowledge about our future and past, it also allows us to 'inhabit' the past and future, to feel 'personally connected' to those events we imagine and remember. They support this claim with the example of patient M.L., who has semantic knowledge about his past life, but feels '*subjective distance*' (p. 603, original emphasis) from it. Co-occurring with this disconnection of M.L. from his past are his problems with acting responsibly. For example, he appeared not to recognise, or be able to act on, his parental responsibilities.

It seems to me that the best way of accounting for this 'inhabited' aspect of MTT is to conceive it as a fundamentally experiential activity. Without an experiential component it is unclear how MTT differs from semantic memory or logical reasoning about the past and future. It is the sensory and emotional experience involved in MTT that characterizes it. In chapter 6 we will explore why these are jointly sufficient for the perspectival character of MTT, in this chapter we are going to look at the contribution of emotional experience to MTT.

I hope to explain why we should see MTT as a type of narrative understanding. Narrative understanding crucially involves emotion, and emotional experiences result in narratives being inhabited. Specifically, narrative understanding involves emotional cadence: emotional sequences that follow from each other in a predictable and coherent way (Velleman, 2003).

The purpose of this chapter is to elaborate on what narrative understanding is, why MTT is a type of narrative understanding, and why a lack of narrative understanding can explain the behavior of M.L. This chapter also proposes that the ability to metarepresent is not necessary for this explanation.

A more distal objective is to start to put together a positive account of the role of emotions in moral agency. The next chapter will explain why narrative

understanding is important for agency, but will bracket the question of moral agency. Chapter 5 will relate narrative understanding back to moral agency. In this chapter we will also see why my account of narrative understanding poses problems for Prinz. In particular, it will challenge his account that concepts and emotions are fully independent.

The last objective, which runs throughout this chapter and the rest of this thesis is to find a place for emotions in our capacity to be subjectively present in virtual situations. That is, to fill the gap that G&K introduced when they failed to spell out a role for emotions in their account of agency. In doing this I show how Prinz's theory of emotions helps us explain why emotions do matter for acting for moral reasons.

2. The lay of the virtual land

First it is important to get clear on exactly what counts as MTT. This concept is vague and evolving, but I'll be using a fairly permissive description used by Suddendorf and Corballis (2007). This account of MTT is important for the claim that to explain the importance of MTT to moral agency, what we are actually doing is explaining why our narrative abilities are crucial to moral agency.

Suddendorf and Corballis suggest that mental time travel refers "to the faculty that allows humans to mentally project themselves backwards in time to re-live, or forwards in time to pre-live, events" (p. 299). MTT is understood to depend on the episodic memory system, which is involved in false, and not just veridical, memory. So, the dependence of MTT on episodic memory means that MTT need not be defined by whether the episode is veridical. Instead it can be defined by whether our mental travelling involves the phenomenological experience of ourself travelling to a different time or place (or, at least some sense in which we feel that we are travelling even if there is also a sense of knowing that we are in fact here and now). The character of this phenomenology will be expanded on later in this chapter, as well as in chapter 6.

This account of MTT suits my purposes here, because it focuses on what is common to both memory and imagination, and my claim about the role of memory and imagination in moral agency also focuses on what they hold in common. Both participate in our narrative understanding concerning who one is and what would make sense for one to do. Remembering my past is not just about making sense of it, but making sense of myself, which is used to decide what I should do in the future. Imagining the future, too, can be used to make sense of what kind of person one is, for example, by registering one's emotional reactions to the imagined events.

So it is precisely what imagination and memory share that is of interest to me here: MTT is important as an ability that allows one to understand oneself, and learning about ourselves directs our actions, and hence our future. Non-veridical MTT concerning the past can perform this function. As explained later, because what is important in MTT is that the emotional cadence expresses something of what we value, the veridicality of the memory doesn't necessarily matter.

What follows from this is that even in cases where we (re)construct our past from information we learn through others or our environment, we may be said to be mentally travelling, as long as it has the right phenomenology. So, Samina, a person with amnesia who looks at childhood photos of herself and listens to her parents recount her childhood adventures, may be said to be mentally time travelling if she uses this information to 'inhabit' a past where these things happen.

A problem might occur if we notice that Samina can also inhabit a childhood that has no factual bases. Note that, while I will later relate narrative understanding, and mental time travel, to agency, whether or not it is veridical, a constraint can be added. I am committed to both narrative understanding about fictional events and non-veridical mental time travel being informative through what they contribute to the formation of our values, see chapter 5. So, for example, I may learn something about the world and what I value if I imagine myself as Stalin, or imagine that I had an impoverished childhood when I didn't. In this way, I may learn something, in particular how I feel and so what matters to me, through how I experience narratives. However, this is consistent with a requirement that if

we are relying on our narrative understanding or MTT to provide us with factual understanding, then it ought to be relatively accurate (see Schechtman, 2007). So, for example, a false belief that we are in Soviet Russia, based on a false memory, would impede my current decision-making. My account need not be committed to claiming that an inaccurate experiential autobiography, that is taken as factually accurate, is conducive to agency.

Another problem that arises is that while we may be able to conceptualise that a person with amnesia is able to mentally time travel, this may only be because we haven't taken into consideration what seems to be an intimate link between MTT and episodic memory (e.g. as noted by G&K, 2010 and Suddendorf & Corballis, 2007). It seems plausible that we need a system that can record past experiences so that we can entertain and inhabit counter-factual possibilities of what could have happened, or might happen. That is, the neurological realizers of fictional MTT may be the same as, or overlap with, the neurological realizers of actual remembered experience. If that is so, then my example of Samina could be leading us to wrong conclusions: people who have amnesia due to the obliteration of their episodic memory systems may find it impossible to mentally time travel.

To accommodate this, I note that amnesia is not a simple affair. A person can have gaps in their memory without having lost all their ability to re-experience the past. Not only can one have blackout due to the heavy consumption of alcohol and other narcotics, but one can have selective loss of memory in both anterograde and retrograde amnesia. In anterograde amnesia, people lose the ability to lay down new memories after a brain injury, but have memories from prior to their injury. In retrograde amnesia, where people lose access to their past prior to their brain injury, it is often only access to certain parts of the past that are lost. In all these cases, some memory is lost, while the capacity for backward time travel may be retained. Finally, considering that it appears that imagination and memory are closely interlinked there is reason to suspect that some preservation of the neurological realisers of the episodic memory system will generally co-occur with preservation of imagination. So, while it seems right that a particular episode of past-directed MTT need not be a memory to count as MTT, this is consistent with MTT being enabled by certain parts of the episodic memory system.

With this in mind, we can still define MTT as our ability to imagine and experience our past and future, regardless of whether the imagining and experiencing of the past is directly caused by our past experience.

I aim to explain MTT through our capacity for narrative understanding. MTT, here, differs from narrative understanding in that it is to do with our personal histories and plans. Narrative understanding, in contrast, is more general, in that there is a sense in which it may not involve us. While we are always prereflectively present when we understand something narratively, the sequence of events may be nothing to do with our daily lives. Think about being immersed in a novel. In a novel you are directed towards characters and events that you have neither encountered nor believe you ever will. MTT, I hope to show, involves a type of narrative understanding: it is understanding a narrative about you or your life.

In this chapter I will develop the view that narrative understanding incorporates an emotional cadence. I argue that emotional cadence is a crucial constituent of having a perspective on events that aren't currently before us. Narrative understanding on this account involves being able to understand narratives through having an embodied and evaluative perspective on a sequence of events.

My view therefore contrasts with that of G&K who argue that it is MTT, rather than narrative understanding, that is necessary for agency. G&K use Bratman, Velleman & Korsgaard to build their view of moral agency, all of whom, at least sometimes, make our explicit understanding of our selves crucial to agency (e.g. Bratman, 2000, Korsgaard, 2006, & Velleman, 1992). G&K take their account to be consistent with Bratman's account of agency, where we are agents because "we conceive of ourselves as agents who persist over time and so we construct and commit ourselves to future directed plans" (2010, p. 601). G&K, like Bratman, make our understanding of our selves as psychological beings that exist through time, a part of their explanation for agency. G&K think that we use MTT to increase our self-knowledge:

> *Planning requires a capacity to imaginatively project oneself into the future; this in turn requires both a sense of oneself as the very same individual who will inhabit that future (autonoetic awareness), and also the kind of detailed self-knowledge that is supported by autobiographical memory.* (ibid.)

For G&K, our knowledge of our particular histories and particular projects and goals is needed to inform our plans.

G&K use patient M.L. as an example of how agency is impaired when MTT is impaired. Patient M.L. has damage to his ventromedial prefrontal cortex (vmPFC), which co-occurs with his problems with MTT. M.L. could only recognise whether events were from his past by his sense of familiarity with them. He was unable to re-experience them and felt subjectively distant from them. These problems with MTT appear to explain his poor decision-making capacity and irresponsible behaviour, which led him to require supervision. G&K take this as a sign that MTT is necessary for moral agency. So G&K build a theory of agency where imagining our own lives is central, through combining philosophical theories with lesion studies. While they bring important philosophical theories and evidence to the debate, I reconsider what aspects of these contributions are relevant to agency.

I now turn to narrative understanding. It is from this starting point that I motivate the view that emotions, as embodied appraisals, are important for the perspectival quality of narrative understanding, and hence MTT too.

3. Embodied narrative understanding

3.1. Velleman's emotional narratives

We saw in the last chapter that G&K want to include emotional processes as one important factor in moral agency. I want to argue that emotions play a role in agency because they are constitutive of narrative understanding. But why would one think that narrative understanding involves emotion? Velleman (2003) starts motivating this proposal with the idea that narrative explanations imbue us with a special type of understanding. Think about the function of fables. When we hear a

fable, it allows us to get a certain grasp on this world. It doesn't seem that the narrative structure of the fable is redundant to the explanation; we don't think a paragraph explaining the moral of the fable provides us with the same understanding as actually hearing a fable. With narratives in general, it seems that we gain something more from them than we would if those narratives were recast in a non-narrative form.

But why should this be? What is the difference between a dry description and a juicy narrative? Velleman suggests that we can discover the way in which narrative contributes to understanding by giving an account of how a description of events becomes a narrative. The key difference, he suggests, is that a narrative is able to initiate and resolve a flow of emotions in its audience. So, it is the capacity of a narrative to "initiate and resolve an emotional cadence" (2003, p. 18) that enables a narrative understanding to engender a particular form of understanding.

It is worth remembering that I will be using the term 'narrative understanding' to refer to a mode of engagement, so that understanding something narratively is a relation between the events being understood and the agent's mode of engagement with those events. Narratives are the object of such engagement.

Velleman goes on to argue that emotional cadence is necessary to explain a crucial feature that narratives enable: that they have plots with a beginning, middle and end. It is noteworthy that a sequence of events doesn't have a beginning, middle and end. A purely descriptive account of history, for example, has no non-arbitrary place to start or end. By contrast, an emotional arc, which follows a familiar pattern of feeling, is responsible for understanding a plot. So it is through the natural cadence of emotions we can experience a sequence as having a beginning, middle and end.

Emotions wax, wane and generally transform in a predictable way in life, and these patterns are stored in "experiential, proprioceptive and kinaesthetic memory" (2003, p. 19). So in narrative, the flux of emotions we have in life is simulated.

We can make sense of this if we recall the theory of emotion we have borrowed from Prinz, that they are embodied appraisals. When we feel an

emotion, what we are experiencing is an embodied sense of our situation. Further, such a sense contains a sense of what actions are possible. So, emotions come with a kind of coupling to our environment, because they are part of how we understand our environment and systematically alter the ways we interact with it. If we take the world to behave somewhat predictably, then we can see how such embodied cadence arises in life: there is a reciprocity between our emotions and our situation, such that we (semi) systematically react to the world, and the world systematically pushes back on us, which in turn (semi) systematically impacts us, and so on.

Importantly, as Velleman observes, some emotions will typically initiate narrative sequences and some generally provide resolution. Grief "can resolve an emotion sequence but rarely initiates one", while fear "can initiate or continue an emotional sequence but it cannot resolve one". Emotions can also "register the impact of a prior emotion" (2003, p. 15). We feel grief if we have lost someone or something that we love.

Consider the fable of the fox who loses his tail:

> *It happened that a Fox caught its tail in a trap, and in struggling to release himself lost all of it but the stump. At first he was ashamed to show himself among his fellow foxes. But at last he determined to put a bolder face upon his misfortune, and summoned all the foxes to a general meeting to consider a proposal which he had to place before them. When they had assembled together the Fox proposed that they should all do away with their tails. He pointed out how inconvenient a tail was when they were pursued by their enemies, the dogs; how much it was in the way when they desired to sit down and hold a friendly conversation with one another. He failed to see any advantage in carrying about such a useless encumbrance. "That is all very well," said one of the older foxes; "but I do not think you would have recommended us to dispense with our chief ornament if you had not happened to lose it yourself." --- Distrust interested advice.*[32]

The story opens with a fearful scene: an animal is stuck. This initiates an action for the main character: struggle. This struggle leads to an emotion of shame, which the audience is expected to empathise with. In fear the fox acts with haste and this resulted with an unfavourable outcome and the sense of a loss of status. Similarly to fear, this shame initiates actions that try to resolve this emotion. Unfortunately

[32] See: http://www.taleswithmorals.com/aesop-fable-the-fox-without-a-tail.htm

for the fox, this action also involves a type of malicious intent, which results from the feeling of embarrassment from the fox's perspective. However it, also a brings a sense of relief for the audience as the sharp minds of the older foxes allows the audience's apprehension, brought on by the potential of a cunning fox tricking others, to subside. This apprehension points both backwards at information the audience has previous learned but the old foxes aren't aware of, and forward, towards the possibility of unpleasant consequences. The tension caused by dramatic irony, in combination with the absurdity of foolish fox's behavior, also evokes amusement in the audience.

Note that the audience is taken on an emotional journey both through empathising with the characters and by taking on an external perspective[33]. The story is strung together as *one* emotional journey because each emotion leads to an outcome and an emotion that is partially determined by the past emotion: fear causes hasty action that leads to embarrassing or shameful feelings which leads to further hasty action in an attempt to resolve these feelings.

The enmeshing of each emotion with the events and emotions that came before it and follow it, allows our narrative understanding to be coherent through time. This coherence is felt as an embodied emotional journey.

Here Velleman relies on DeLancey's (2002) version of an affect program theory. This theory suggests that psychologists and philosophers should delineate emotions by:

1. The type of circumstance that triggers them;
2. The physiological changes that follow the trigger;
3. How they dispose us to behave;
4. What conditions lead them to decrease in intensity, and the thoughts that they tend to result in.

Velleman's suggestion then is that we understand narratives when our engagement with a series of events allows us to cycle from steps 1 to 4 in a predictable way. By stage 4, we are also back to stage 1 of another emotion. And since emotions involve physiological changes, when narratives take us through the

[33] Goldie calls this slipping between perspectives 'free indirect flow'. I will return to this concept later.

predictable patterns of emotion, the narrative makes sense to us experientially. Narratives get under our skin, and we understand them in our bones[34].

While I don't want to defend the notion of affect programs in it standard use[35], there are ways to re-evaluate and reconfigure the characteristics above in relations to understanding emotions as embodied appraisals. Embodied appraisals, as containing both action affordances and being part of our perception of events, are characterised by what the affect program theory understands as stages. Step 1 is an explanation for how embodied appraisals come about: through the registering of a situation. Step 2 & 3 is exactly what is embodied about embodied explanations, although we do not see these as discrete steps, on an embodied appraisal theory, but as two aspects of the same phenomena. Step 4 seems a reasonable inference if we take emotions to be embodied appraisals, that is, actually saying something about what our situation is. On this interpretation, if our situation changes, then our emotions change. And, just like we take our senses to influence (or be integrated) with what we think, because we think they have some epistemic value, so, if emotions have some epistemic value, we should take it that they will influence (or be integrated) with what we think.

Combined with the understanding that such embodied appraisals also involve action affordances, we come to understand why a theory of embodied appraisals implies the existence of emotional cadence. Our environment and emotions are coupled together through our actions varying somewhat predictably based on our emotions, and our situation varying somewhat predictably with our actions. In this way, we can get the cycling through the steps that I mentioned above. Step 3 and 4 contribute, in combination with a semi-predictable environment, to what emotion we feel next.

Velleman also thinks that the involvement of emotions in narrative understanding explains why narrative understanding involves a unique grasp of

---

[34] Strictly speaking, narratives are not understood in our bones. If they are understood emotionally, and we have an embodied account of emotion, then they are understood through sensing our organs, the rhythm of our breath, the position of our muscles and joints, and the accompanying feeling of comfort and discomfort (Barrett & Bar, 2009).
[35] See review of the idea by Colombetti (2014) who critiques the idea that emotions are hard-wired in the way suggested by proponents of affect program theories.

events as a coherent whole. That is, the emotion we are left with at a story's end expresses a stance on the story in its entirety. Because the emotional resolution of a story often points back, and subsumes, the previous emotions, we are left with a sense of what the story means for us.

For example, in the fable of the fox that loses their tail, we may be left with at least two emotions: empathy with the embarrassed fox and relief that no-one got tricked. This gives us a stance on the story that is, unsurprisingly for a fable, an understanding of an important life lesson. For it tells us that we will be shamed for acting like the naughty fox, but also teaches us the value of being shrewd in assessing whom to believe.

Plot and coherence are therefore put forth as characteristics that allow us to understand narratives in a different way to just a description of a chain of events. Our emotional engagement explains how it is possible for narrative to have these defining features.

We may object that a description of a causally-connected sequence of events can be coherent via its inferential commitments and connotations. What is important to remember is that Velleman is explaining the sense in which a story has coherence that cannot be translated to description. So, while understanding pure text may involve understanding coherences in some ways,
Velleman is trying to explain the coherence peculiar to narrative understanding.

It doesn't seem true, though, that all the narratives we engage with are fully coherent, with a nice beginning, middle and end. Velleman's response to this is that some of the things we think of as narratives: novels, oral story-telling, film etc., may employ narrative as part of their tools, but are not always reducible to narrative. Velleman argues that we if read a book that leaves us hanging, is firmly irresolute, then that is because it is no longer employing narrative techniques. Nonetheless, we could make a couple of other responses to this observation. One is that narrative understanding consists of the knitting together of emotions so that, in a significant proportion of a story, each emotion develops from the preceding one in a way that is familiar to us. This loosens the need for the final emotion in a story to resolve the previous emotions. Additionally, artistic narratives may diverge from some of the structural regularities in our everyday narratives. This is

similar to another suggestion Velleman makes: artistic narratives often toy with the usual way that we would understand events narratively to create discordance.

Furthermore, while there is a sense in which narratives are those things we listen to, watch, read, it should be remembered that 'narrative understanding' is a much broader category, and the focus of what we are trying to explain. Narratives, in the sense of those things delivered by others in prose, pictures, or on screen, are good at engendering narrative understanding in us in virtue of the techniques used. Narrative understanding remains the focus: that special way of apprehending a string of events emotionally.

Velleman makes a convincing case that emotions are one way our understanding of a description of events becomes narrative, although some might not think he has done enough to show emotions are necessary for narrative understanding. I want to set the necessity claim aside in lieu of an alternative process that explains the existence of plot and a unique of sense of coherence gained through narrative understanding.

It is evident that an embodied, affective component of narrative understanding allows one to explain the features of narrative that Velleman is interested in. Without a perspective embedded (as-if) in a string of events[36], we do not have beginnings and ends. Beginnings and ends are not inherent to chains of events, because there is no reason why any particular part of a chain is a start of something or its conclusion. Devoid of a point of view from somewhere, any point preceding or following would do equally well. Similarly, Velleman makes a compelling case that emotion is one means for providing coherence, both for the narrative moving along in a way that makes sense and for the holistic understanding one associates with narrative.

However, as will become clear throughout this thesis, Velleman's theory of narrative understanding is also compelling because it fits together with, and makes sense of, many theories of (moral) agency. Furthermore, in this chapter and chapter 5, we will be presented with a good deal of empirical evidence to support

[36] I will come back to this point about having a perspective in a string of events when discussing Goldie's view of narrative.

the claim that affect is one crucial constituent in having a perspective on past, actual and fictional sequences of events.

So, while leaving open the possibility that we can explain plot and coherence in narratives by means other than emotion, in the rest of the paper, when I refer to 'narrative understanding', I mean the type of understanding that involves emotional cadence.

3.2. Emotions as perspective in narrative

Velleman's enterprise starts with the realisation that narratives have a certain type of meaning for us. He analyses this special type of meaning by the way that narratives embed us within the situations they describe. Velleman describes narratives as meaningful to us because the affect programmes they trigger allow us to inhabit stories viscerally and kinaesthetically. That is what is meant by 'narrative understanding'. I think this interlocks with an observation made by Goldie, and a claim I want to make about the embodied perspectival nature of narrative understanding.

Goldie (2012) has claimed that a distinguishing feature of understanding narrative is that it involves a perspective or several perspectives. For him, an external perspective is necessary for understanding a narrative. The "shaping, organising, and colouring [of the narrative] is informed by something that is at the heart of my account of narrative: the narrator's perspective or point of view from which the events are narrated" (p. 11).

The external perspective is able to provide "the three characteristics of narrative: coherence, meaning, and evaluative and emotional import" (p. 40). This perspective organises a sequence of events into a coherent whole, and provides a meaningful interpretation and an evaluative or emotional response. In MTT the thinker who thinks through the sequence of thoughts that make the narrative is also the external perspective.

I want to argue that a slight reformulation of Goldie's proposal is needed: emotion is one form that perspective takes. As I will outline below, this is because emotions are involved in our feeling of how we extend into space, how our

situation bears on us, and what actions are available to us. I will expand on these claims in chapter 6. This account of emotion as a form of perspective helps us understand how narratives are inhabited. That is, it allows us to grasp more fully how emotions can do the work within narrative understanding that Velleman claims that they do.

When we understand that emotions co-emerge with perspective we realise that the characteristics Goldie identifies with narrative understanding are interdependent. We should not think that perspective is the cause of coherence, meaning and evaluative and emotional import. Rather, being a creature for which there is coherence, meaning, and evaluative and emotional import is part of what it means to be a creature with a perspective. This is not an argument against Goldie, but a proposal of the most insightful way to understand his account. At the least, it is a good way of summarising mine. My proposal for what moral agency consists in includes an understanding of how I think perspective and emotion co-emerge, and an explanation of why the ability to understand events as meaningful wholes is constitutive of being a creature for which there is a moral outlook.

In agreement with Velleman, I think the best way to talk of the role of emotions in narrative understanding is as placing a person within a chain of events. While Velleman does a good job of highlighting the way that emotions bring a particular type of temporal continuity to a story, he also touches on something I expand on below: that emotions not only allow us to live through a story temporally, but can place us within the scenes of a story. This is also what Goldie is getting at through his explanation of narrative understanding as perspectival. Words, I suggest, bring an imagined world before us by giving rise to a pattern of physiological responses. This point depends on a particular theory of emotions, that they are 'embodied appraisals', because it is the embodied character of emotions that enables us to explain how narrative understanding involves the phenomenology of inhabiting

counterfactual scenarios. I will be assuming this theory of emotions throughout this chapter and beyond[37].

If emotions are 'embodied appraisals' (Prinz, 2004) then they express a sense of one's situation, what matters to oneself in that situation, and a range or affordances. These characteristics of emotions situate a creature.

Although the particular term 'embodied appraisal' refers to a theory of emotion by Jesse Prinz, a similar idea has been expressed by others (Ahmed, 2004; Colombetti, 2014; Ratcliffe, 2005). Emotions, in these theories, are simultaneously physiological changes and a sense of how the world is for an organism. Further, Prinz (2004) names these physiological changes 'signals' to indicate that they function both as a perception of how one is related to the environment and as action tendencies. So, feeling excited when I see a piece of chocolate is both a perception that I am in the presence of something that, if I can get access to it, would be beneficial, and it prepares me for acting in certain ways: in this case it might prepare me to pick up the piece of chocolate and put it in my mouth[38].

As such we can conceptualise emotions as having a bipolar structure. Since emotions are perceptions of how the world is related to an organism, an emotion expresses that 'I am here' on one end of the bipole, that place that is being affected, and that which does the affecting is 'over there', external to me (Ahmed, 2004)[39]. Because emotions are physiological changes that are felt, 'I am here' is not just a

---

[37] There are many other ways of conceptualising emotions. We can be a cognitivist, that is, see emotions as caused, or constituted by, propositional attitudes (e.g. Lazarus, 1991), and that such cognitive components are disembodied. The opposing view is to argue that emotions are the feelings of bodily changes (e.g. James, 1884). Yet, as we saw in my first chapter, Prinz has presented some good reasons for understanding emotions as embodied appraisals, which differs by not depending on propositional attitudes, but maintaining that our bodily feelings are appraisals. As he noted, while emotions co-vary with our understanding of a situation, there are experiential and empirical reasons for thinking that they are embodied, and do not require explicit thought to be triggered. Furthermore, this theory of emotions has the explanatory value of making what initially appear to be disparate notions of moral agency consistent.

[38] In this particular respect, any appraisal theory could play this function. However, the embodied nature of emotions is important for other dimensions of the perspectival nature of emotions, which I shall mention shortly.

[39] Disgust may complicate this: we might think of it as the appraisal that something external to me is contaminating me, or within me, that the boundary between internal and external has been violated in some way (Ahmed, 2004).

conceptual truth, but feeling anchors a creature in the physical world experientially. I am where the emotions are felt. Perspective co-emerges with an ability to locate ourselves in the world.

Further, emotions are perceptions of what external things matter, and how they matter. That lion running towards me matters, and it matters because it can kill me. Perspective co-emerges with the ability to grasp that a world bears on me, and how it bears on me[40].

Finally, emotions involve taking a stance on the world in so far as they "allow us to literally perceive that situations afford a range of behaviour responses" (Prinz, 2004, p. 228)[41]. So if we see a lion and feel fear our world is experienced as consisting of several different routes to run through, a grass to fall and freeze on, or places to hide behind. Perspective co-emerges with the ability to grasp how I can act on the world[42]. In chapter 6, I will return to the subject of what it means to have a perspective, and how perspective arises through examining the sense in which affect and sensory experience are interdependent and, in their interaction, contribute to perspective.

There are two ways in which narrative understanding can have a perspectival quality, and both may occur when engaging with the same narrative. When we understand narratives, we can place ourselves within the narrative as a character, or we can have an external perspective watching what happens.

---

[40] Slaby & Stephan (2008) have a similar view of affectivity as self-disclosing through "[making] manifest what is currently of relevance to us" (p. 508). Bemudez (2001) also ties nonconceptual self-consciousness in both with our sense of our body, the way it delineates self from world, and the way it is incorporated into perception as a sense of what the world affords.

[41] For a similar account of emotions see Slaby (2012) where he understands emotions as the type of self-awareness that comes with 'a sense of ability'. Although she doesn't mention affect, Hurley's (1997) argues that the perspectival quality nonconceptual self-consciousness emerges with an ability to keep track of the relation between our actions and what is perceived.

[42] Note that although I can express these characteristics of perspective and emotion separately, I am not claiming that, within an occurrence of emotion, these characteristics come apart.

The view that emotions co-emerge with perspective makes sense of Goldie's observation that the way our emotions relate to the narrative coincides with whether we have an external perspective on, or internal perspective as if from within, a story (2012, p.11). He notes that if one's emotions match those of a character within a narrative then our perspective is internal to the narrative. That is, we take the character's perspective. If emotion matches that of the external narrator, then the perspective is located 'outside' the story looking in.

MTT, as narratives about our own lives, takes both these forms. We can create and understand narratives of our past and possible life while taking the perspective of someone external to the events. In which case our emotional reactions are directed to the characters, including ourselves, and events, and those emotions contain something of our assessment of what the events and actions within the narrative mean for us. However, there are times where we become ourselves within an imaginary (or remembered) realm. At such times the emotions are those of the character acting out the story. Those emotions situate us as an actor within the narrative, and are involved in our assessment of the imagined situation.

So, if I am remembering that time I argued with my sibling, the experience of anger is more likely to place me within the story, as that thing being attacked, while the feeling of guilt is more likely to place me external to the event, as that thing that, in retrospect, feels bad for shouting.

In the case of viewing a narrative from the inside it might seem that there is no *narrator* perspective, which Goldie claims is necessary for narrative. Yet in this case the internal perspective is able to play the same role as he gives the narrator perspective. It is involved in "shaping, organising, and colouring" events from within. Further, at the least it seems that narratives often switch between external and internal perspectives. So, if we think that a narrator is external to the characters within the story, in any particular self-narrative there are likely to be some points within it that contain a narrator's perspective.

It is also a possibility that both perspectives can be taken simultaneously. Goldie talks of 'free indirect style' as a characteristic of many narratives (2012, p. 32-40). It occurs when it is ambiguous which perspective is being taken, or when

two perspectives seem blurred. The theory that emotion is a form of perspective is also consistent with the possibility of ambiguous and mixed perspectives – mixed or ambiguous emotions constitute co-occurring or ambiguous perspectives.

We can also understand Velleman's idea that it is emotion that provides us with a gestalt sense of the entire narrative in this light. The emotional state we are left in at the end of a narrative encapsulates a perspective on the narrative, an embodied sense of what the completed narrative means for us.

So, we should see Velleman's idea that the essential work emotion does within narrative understanding is to re-awaken patterns stored in "experiential, proprioceptive and kinaesthetic memory" (2003, p. 19) as another way of spelling out Goldie's view that perspective is a distinguishing feature of narrative understanding. Emotions work to bring us within the situation that the narrative lays out, and in doing so come together with a perspective in and/or on the story[43].

At this point, we might be wondering how extensive I am taking 'narrative understanding' to be. Is all experience narrative or only some experiences? Under my view, narrative understanding can be a matter of degree where the extent of it can vary under several dimensions. Much of human experience falls, to some extent, under the category of "narrative understanding". Experience is narrative depending on intensity of its affective quality, and the extent that the affective experience is coherent, where, as Velleman suggests, coherence can be understand as occurring both through time, and when we take an overall perspective on a sequence of events.

Further, as we shall see in chapter 5, narrative understanding depends on some kind of recombinable system – which is the capacity to have some kind of understanding where parts can be used in various ways. (I will expand on this explanation later). In the human case, we often have language – a type of recombinable system that we can use to reflect. Thus the type of narrative understanding many humans have is special – they experience things narratively not just in the sense that there is an emotional cadence to their understanding of

[43] None of this means that emotion is the only process in perspective-formation.

events, but in the sense that they can reflect on and revise their (affective) understanding of those events.

A recombinable system produces another dimension through which we can understand the extent to which some experience is narrative. Narrative understanding, as a gestalt, can be the coming together or more or less parts. The combination of just a few parts does not encompass the richness of coherence achievable in a complex system. Because language, in the human case, is a complex recombinable system, the type of gestalt that is emerges with are also rich and complex. Human narrative experience therefore also varies as a dimension in which our affective experience can be more or less richly narrative, because the type of coherence these experience have can be more or less complex and rich.

To the extent that that some types of deliberative thinking might be less strongly affective (but still experienced in some way), it is experienced less narratively. And to the extent that coherence of experience breaks down, perhaps in psychosis, experience is less narrative. An experience is narrative in a less sophisticated way when it is structured by a simpler recombinable system than language. An experience is not narrative at all if it is not structured by a recombinable system.

3.3. Scientific support for emotion in perspective and narrative

There is good empirical support for the claim that emotions are involved in perspective formation within imagined and remembered narratives.

Markowitsch and Staniloiu (2011) review evidence that emotions, among other things, are involved in diachronic self-awareness. Particularly they are concerned with what is involved in 'episodic-autobiographic memory' (EAM). That is, memory that involves knowledge and experience of self as existing through time. Reviewing the neurological literature shows that emotional hubs, such as the amygdala are active during EAM. So, they argue that,

> *The reliving of the subjective experiences from the encoding context is usually intimately linked to an emotional evaluation of the significance of these past experiences for oneself and with respect to one's own position in his [sic] social and biological environment. This emotional evaluation may*

> *in turn shape someone's motivation for planning for the future and engaging in future acts.* (p. 19.)

However, other psychologists they cite make a bolder claim, arguing that emotion is "intrinsic" to EAM (p. 20). This is supported by evidence that brain circuitry associated with providing the emotional, biological and social significance of information is involved in EAM. Circuits thought to be involved in these tasks, the Papez circuit and basolateral limbic loop (which includes the amygdala), are activated when people engage in EAM. More circumstantially, the ability to engage in EAM, both within an individual's lifetime and through evolutionary development, are thought to be reflected in structural and functional changes that include circuits involved in emotion. Similarly, the degeneration of von Economo neurons (VEN) is associated with both decreased emotional awareness and changes in self-consciousness. Levine et al. (2004) also cite several imaging studies where "self-generated autobiographical material show activations in ventral limbic and paralimbic regions associated with emotion…even when emotional memories are not explicitly requested"[44] (p. 1641).

Markowitsch & Staniloiu also cite evidence that the activation of the amygdala increases if one experiences oneself in the midst of the situation rather than observing from outside. I note that even an external perspective is in some sense inhabited. If we are external to a narrative there is still a sense that we have/take a perspective on it, since we place ourselves outside of the narrative looking in. Nonetheless, our sense of perspective presumably comes in degrees, a sense that we are in the midst of a situation, rather than observing it, may increase this feeling.

Finally, the view that narrative is importantly emotional is congruent with neurolinguistic evidence. Brain imaging has revealed that understanding

---

[44] This evidence falls in the context of a theory by the neuroscientist Antonio Damasio (1999). He posits that emotion and implicit self-awareness of oneself in the present (what he calls "core consciousness") share largely overlapping neural structures. The function of these structures, he believes, is to represent the body and its potential actions, and how it is related to the world. He reaches this conclusion mainly through his observations of different neurological conditions, where lack of core consciousness and emotions coincide.

narratives involves the amygdala and limbic system to a greater degree than word and sentence comprehension. These are brain regions associated with emotion (Xu et al., 2005).

The view of emotion I have outlined above chimes with these data and claims from neuropsychology. If emotions are embodied appraisals then the feeling of an emotion is the feeling of our body as it relates to the world. It is no surprise then, that emotion is involved in consciousness of the self, and therefore in self-referential processes.

4. Narrative understanding & mental time travel

4.1. Mental time travel as a type of self-narrative

We started off defining MTT as the ability to subjectively (re)experience our past and future, but now it looks like it just might be a type of narrative understanding. Narrative understanding, after all, requires us inhabiting a string of events either by our experienced emotional cadence co-emerging with us surveying those events from the outside, or through our emotional journey placing us within the events (or any mixture of the two). Narrative understanding is essentially perspectival and inhabited, and involves understanding events through time, both characteristics we take to define MTT. One crucial component that appears to be present in our understanding of MTT but missing from this portrayal of narrative understanding is sensory imagery. Yet I think this distinction fails, as I explain in the chapter 6 – for sensory experience is also necessary for narrative understanding.

I think we can make sense of MTT by understanding it as a type of narrative understanding, one which concerns our selves and our lives. Once we do this, we can understand why MTT motivates our action.

If we go down this route, we can also explain why MTT has an embodied perspectival quality. In part, it is because of its emotional character. I propose this perspectival quality is the sense in which one can 'inhabit' MTT. MTT has this phenomenology because it engages our emotions. So, in the case of M.L., who

doesn't feel subjectively connected to what has happened to him in the past, it is likely that this is due to him not feeling emotionally connected to those events. The reason relates to Velleman's explanation of narrative understanding: narrative understanding relies on an emotional continuity through time. In the context of typical instances of MTT, this means that the story that leads from our past to our current self, or from our current self to our future self, involves some emotional continuity such that we feel connected to our time travelling selves. MTT involves an emotional cadence in the sense that each emotion is coherent with the emotion it follows from and leads to. Because emotions and perspective co-emerge, emotional continuity can involve the sense that the same embodied being is involved in the past and the future. While emotional cadence is necessary for plot, it can also result in the continuity of a character's perspective. In what we normally take to be standard cases of MTT, we often identify the character's perspective as our own.

Returning to our scenario at the start of this chapter, concerning Samina the (partial) amnesiac who recreates her childhood based on experiences she has no direct access to, we can also explain why this counts as narrative understanding. While her narrative is not caused by the events that happened but is constructed, what matters is that the thing that she has constructed includes an emotional cadence that creates continuity between her childhood self and her current self. Not only does she mentally time travel, but her MTT has a narrative structure. However, for Samina's re-construction of her past to give her appropriate factual information about her *own* life, we should note, as before, that it should be relatively accurate.

It is Samina's ability to construct narratives that enables her to mentally time travel. She can imagine what would have been a coherent way for her younger self to have felt and can relate this to how she feels now. Further, for Samina to be able to mentally time travel it is enough that she can create narratives about a possible way her life has been or will be; it does not need to involve factual information. For us all, this is sometimes the case, as when we imagine our futures. So narrative understanding is sufficient for us to engage in MTT.

This interpretation of the role of emotion in MTT, as suggested at the start, depends on identifying what memory and imagination hold in common. While much of the above scientific evidence concerns memory, it is likely that we can extrapolate it to imagining the future too, as they share a similar phenomenology. If we take emotion to play an important part in the ability for us to inhabit our past, and we also need an explanation for how it is possible to inhabit the future, then it seems legitimate to put forward emotion as an explanation for how this is also possible. So, for example, my feeling of nervousness before a session with my supervisors is constitutive of my perspective on a likely future situation in which my ego is damaged. My feeling of embarrassment as I imagine my exegesis of Charles Taylor being picked apart is constitutive of a future scenario that I place myself within, where my embarrassment is constitutive of a perspective as-if I am currently being criticised. Here, there is some sense in which emotion places me in an imagined future scenario. Embarrassment, here, is part of the construction of an imagined scenario where my faults are on show.

The importance of affect in narrative understanding provides a way to explain why people can act on what they discover through narrative imagining. As emotions are both an evaluation of one's relationship to the environment and present a range of affordances for action, they are involved in our capacity to act on what matters to us.

On this account of narrative understanding, all narrative understanding is pre-reflectively self-involving. Even when we read a story, or listen to our friends recount their latest adventure, we understand it narratively when we have a sense of perspective, constituted by our emotional cadence. Such narrative engagement is *self*-narrative in a way: all narrative understanding depends on the involvement of our perspective on it, it depends on how we, as embodied subjects, relate to it. In this particular sense, all narrative understanding is self-narrative. Such narrative understanding does not depend on us thinking about ourselves as the topic of the narrative.

Narrative understanding occurs also when we engage in MTT and when we conceptualise ourselves diachronically. When someone engages in MTT they understand narratives about their own lives. Note we may be prereflectively present here too. We may remember events that happen to us, and experience them as happening to us, without really thinking of ourselves as psychological beings. But then we may also tell metarepresentational stories: we may explicitly understand ourselves as representational things. For example, we may understand our beliefs as beliefs. In metarepresentational stories, this enables us to conceptualise ourselves as diachronic psychological beings with both the experience and knowledge of the character known as 'me'. Experiencing the flux of emotions as we mentally time travel is partly constitutive of experiencing ourselves as a continuing character (that character that is me) through time; and the explicit story can also contain references to our history, relationships, the future plans, the intentions, hopes and fears. Think of Samina. When she understands herself narratively she could be representing herself as a psychological being as well as experiencing herself diachronically.

G&K take this last type of self-narrative to be a fundamental capacity for making decisions: not only do we experience ourselves as subjectively present in MTT, but this prereflective awareness is integral to building metarepresentational stories, which allow us to make plans that we are committed to. That is, according the G&K, decision making depends on understanding metarepresentational stories: stories that are explicitly about who we are and our mental states. Such stories are connected to us in the right way, through us inhabiting these stories, and so we are motivated to act on our plans. Now I want to explain why MTT is not necessary to explain the behaviour of M.L. and one other vmPFC patient. Because a lack of narrative understanding, a non-metarepresentational capacity, is sufficient for explaining their behaviour, an inability to metarepresent is not necessary to explain their behaviour. In the next chapter I will expand on this notion that MTT and metarepresentation are both specific instances on narrative understanding.

4.2 Explaining ventromedial patients without metarepresentation

While G&K make MTT crucial to stories about loss of responsible decision-making, I think we can explain M.L.'s behaviour without claiming that his poor decision-making is due to his inability to mentally time travel. If M.L. lacks narrative understanding then he loses the capacity for understanding virtual scenarios in an affective and embodied way, a way that is linked to his capacity to act on his understanding of counterfactual possibilities. His inability to understand events narratively, to feel his way through virtual worlds, underlies his inability to MTT. That is, we can identify a common cause that would both prevent M.L. from mentally time travelling, and behaving responsibly. Both require narrative understanding.

Note that metarepresentation is not necessary for narrative understanding. You can have a sense of an outcome being bad, and that can influence your actions, without explicitly thinking about yourself. The capacity to foresee the total devastation caused by climate change could prompt you to recycle, and thinking about your own life would not be necessary. Much of M.L.'s irresponsible behaviour may be due to his inability to engage in narrative understanding like this, rather than his inability to engage in MTT.

MTT, as relating more directly to your life, may be important too, but that just comes with the ability to understand events narratively. So, for instance, MTT informs your decision not to go on holiday if you envision too much work waiting for you when you get back. Again, you don't *have* to be representing yourself as a psychological being. You have to have certain psychological processes, like believing that having too much work is bad, but you don't have to represent yourself as having those processes. Here, you are thinking about your life, but you are not yet metarepresenting.

Similarly, the case of EVR is congruent with him having a problem with narrative understanding in general, rather than MTT and metarepresentational MTT in particular. Patient EVR, who also has a vmPFC lesion, is hopelessly indecisive and finds it hard to act on decisions he does make. Again, we can see that EVR would have this problem if he lacks narrative understanding, if he lacks the possibility of inhabiting counterfactuals and what they would mean for him.

I am not currently claiming that we have reason to accept my hypothesis over the alternative. This is just to claim that there is a competing hypothesis to the one that G&K propose. In my next chapter I will argue that my hypothesis, that agency is a matter of being a creature capable of narrative understanding, is preferable to their alternative, that agency is a matter of conceiving of oneself as a psychological creature.

What is important here is how wide-ranging our narrative capacities are; how inhabiting counterfactuals occurs frequently without us explicitly considering our own lives. Decision-making incorporates understanding we gain from our narrative understanding of: political, social and economic events; history; the lives of famous people; the television shows, films and plays we watch; the books we read; the stories our friends tell us about their lives and the lives of others they know etc. It is unclear why metarepresentation is essential for the decisions we make using these narratives.

This explanation faces a potential objection. I've made emotions central to the perspectival nature of narrative, and yet these patients have suffered a lesion to their ventromedial prefrontal cortex (vmPFC), not an area of the brain always associated with emotions[45].

As G&K argue, there is good reason to think that the ventromedial prefrontal cortex (vmPFC) plays some kind of role in enabling a sense of self. I think there is reason to think that vmPFC does this by playing a more general role in the capacity to have a sense of perspective on a complex situation or a series of events – whether it is one's own or another's. This is because of how I think we should understand the common denominator in the different activities that the vmPFC contributes to.

First, the vmPFC is implicated in cognitive empathy i.e. the ability to understand how another person understands the world (Walter, 2012). Second, in an analysis of several tasks where the only common theme is to construct a scene,

[45] One thing to be clear about here, is that I am not suggesting that emotion is sufficient for the complex perspective-taking involved in narrative understanding. Not only do I think that sensory experience is also necessary (chapter 6), but (as will become clear in chapter 4 & 5) conceptual processes are too.

but where only some were directed to explicitly self-involving processes[46], the vmPFC turned out to be part of the circuit involved in all conditions (Hassabis & Maguire, 2007). Scene construction here involves "the process of mentally generating and maintaining a complex scene of event"(p. 299). Finally, as G&K note, vmPFC is activated in the recollection of autobiographical memories (Gilboa, 2004). The common denominator of the capacities implicated in these studies is, I suggest, that vmPFC is involved in inhabiting a perspective where complex temporal and/or spatial factors come together, whether it is one's own or another's. If scene construction is not obviously indicative of this, note that we cannot have a scene before us unless there is at least an implicit sense of the self that the scene is before. This will be expanded on more fully in chapter 6.

In line with this suggestion, Roy et al. (2012) argue that the vmPFC is a hub region, or 'a system of systems' that co-ordinates "episodic memory, representation of the affective qualities of sensory events, social cognition, interoceptive signals and evolutionarily conserved affective physiological and behavioural responses" (p. 147). In doing so, it allows an organism to conceptualise itself in context, what that context means for it, and thereby understand the 'affective meaning' of a situation. They used a database of 4,400 studies to pick up patterns in data, and supported their analysis with animal lesion studies. From this it appears that the vmPFC is an integrative centre for affective, motor, and simulation systems. It is action guiding, but particularly in regards to complex situations and the context-specific formation of goals.

I think we can see how this is congruent with the vmPFC patients having a problem with narrative understanding in general, rather than MTT or metarepresentational MTT, in particular. Narrative understanding is a perspectival view of a series of events, where affect is a crucial constituent in enabling an embodied sense of what a current or counterfactual situation is. It basically is the capacity for an embodied understanding of particular, complex situations and their

[46] For instance, tasks that obviously involve the self included remembering ones own past, while general imaginative tasks, navigation and vivid dreaming were counted as tasks that did not involve an self-referring characteristics. The authors of the paper claim that these tasks don't involve self-awareness at all, but this may be due to lack of consideration of the possibility of prereflective self-awareness.

outcomes. It is no surprise then that if you knock out the vmPFC, the perspectival nature of understanding events can be significantly harmed. The vmPFC co-ordinates and integrates affective information with other systems. Remove the vmPFC and you remove a part of the brain that contributes significantly to a perspectival and embodied view on a sequence of actual and non-actual events.

5. Conclusion

We have seen that G&K have got something very importantly right. It is in virtue of certain characteristics of MTT that MTT contributes to our capacity to make responsible decisions, and to act on decisions we do make. There is something important about being able to subjectively inhabit non-occurring events. When you can do this, you can understand how the possible actions you are faced with now will play out, and you have prereflective access to what those consequences will mean for you. Your decisions, reached through consideration of counterfactuals (whether in your life or through fiction), are embodied and action-orientated. Nonetheless, I have argued that this capacity is based in narrative understanding, which is broader than MTT. A competing explanation, therefore, to the one that G&K propose, is that vmPFC patients have problems with decision-making because they have damaged narrative understanding, rather than because they have a particular problem with MTT. Further, when we understand what narrative understanding allows us to do, it seems we need some further motivation for why metarepresentation is needed for good decision-making.

What I want to do now is examine why narrative understanding, not MTT in particular or metarepresentation, is sufficient for agency. Narrative understanding, I argue, is what enables MTT and metarepresentaiton, and moreover narrative understanding is sufficient for us to be a unified self.

In chapter 5 I argue that the combined emotional and conceptual content of narrative understanding is crucial for its role in moral agency, by allowing us to be interpretive creatures that can articulate what is of qualitatively higher value. Finally, in chapter 6, I complete my account of narrative understanding and how it

is involved in moral agency by expanding on how it is possible to inhabit a sequence of events. I argue that it is through the interdependence of sensory and emotional experience. Because such interdependence enables the experience of an embodied presence in narratives, it allows us to act on our narrative understanding, through enabling an embodied identity between our current selves and the narrative sequence. Thus, in narrative understanding, our moral sense and action affordances are bound up together.

Narrative Agency

The Universe is made of stories
not of atoms.
Muriel Rukeyser, 1968

1. Introduction

In the last chapter we saw that narrative understanding competes with metarepresentation and mental time travel (MTT) as an explanation for how we make responsible decisions that we are able to act on. The issue of agency was bracketed because I wanted to first get clear on what narrative understanding is, and why it includes, but is not reducible to, MTT. Here we return to the issue of agency. For the sake of scope limitation, I will continue with the foundational premise behind the claims of Velleman, from whom this theory originates. And that is that an agent is a creature who cares about acting in a way that makes sense. This is the sense in which an agent is a rational being.

However, G&K tie this ability to the capacity for metarepresentation. What is pivotal for them, beyond the experience of ourselves as diachronic, is that we use our self-knowledge, garnered through autobiographical understanding, in order to plan. By claiming this, they make MTT relevant to agency insofar as it relates to our capacity for "causal-psychological" reasoning. Like G&K, I think MTT has some features that are crucial for agency. Unlike them, I think this is part of a broader capacity of narrative understanding. This capacity does not make metarepresentation essential to agency.

To sharpen this contrast, I will explain Velleman's distinction between his causal-psychological and narrative theories of agency. Contrary to Velleman, I argue that you can explain the former in terms of the latter. This provides conceptual reasons for thinking that much of the contribution of MTT and metarepresentation to agency relies on our capacity for narrative understanding. As such, narrative understanding, rather than MTT and metarepresentation, has explanatory priority.

Yet, this does not yet explain how narrative understanding, in the absence of metarepresentation could contribute to agency. So, further, I will explain how narrative understanding could be sufficient for being a coherent agent that can do what would make sense, so a capacity for reasoning about the psychological is not necessary. For an agent's coherence can be created through our narrative understanding forming systematic attitudes towards the world, which does not require metarepresentation.

Once we have this distinction clear we will be able to see what is right and what is wrong about Galen Strawson's attack on narrative theories of selfhood. His attack is prominent in the debate concerning narrative understanding, and therefore a response to his objections is warranted. I will argue that he is right to be sceptical that metarepresentational narratives are a fundamental part of our psychology. Since metarepresentation is not endemic in human thought, making metarepresentation central to agency rules out too many people as agents. Nonetheless, I will argue that he is wrong to conclude that a more minimal narrative understanding is either trivial or unnecessary. Both of my responses to Strawson fall out of the richer explanation of what narrative understanding is that I developed in the last chapter. Since we now have this richer view on the table, we are able to clear up some confusions.

The aim of this chapter is therefore to get clear on what my minimal requirements for agency are, and why. The claim is not only that narrative understanding meets the minimal criteria for agency, but that it allows us to explain many other capacities involved in agency. Note that the claim is not that other capacities don't enhance our agency. I now turn to Velleman, and his two supposedly distinct ways of spelling out the sense in which we can be agents.

2. Narrative understanding & causal-psychological understanding

2.1. Velleman's distinction

To get clear on what the dispute is between my account on agency and the one that G&K suggest, it helps to look at a two apparently distinct accounts of agency

provided by Velleman, which I will do in this section. In section 2.2, I will suggest that narrative understanding underlies our capacity for metarepresentation. However, we may not yet understand how narrative understanding could be sufficient for agency. Representing a self as a self may be necessary to explain how we can get the self-consistency needed for agency. Why narrative understanding is sufficient for this is explained in section 2.3.

In his 2007 *Reply to Catriona Mackenzie* Velleman admits that he has been using two parallel theories of agency without acknowledging that they are different, and without explaining the way they are connected. Initially, Velleman described an "autonomous agent as understanding himself in causal-psychological terms" (2007, p. 284). This is an essentially metarepresentational understanding. It means that I have "the ability to understand what I am doing as done by a creature who I am" (p. 259, 2006). This allows me to, "round on my entire self and wonder, "What is this creature up to?"" (*ibid*., p. 260). This ability allows us be coherent because once we have beliefs about our motives, beliefs and goals, we can choose an action that make sense in light of our psychological character[47]. For example, if Pablo believes that sexism is wrong, and he realises that he has some implicitly sexist attitudes, then he may be motivated to find ways to change his implicit attitudes to better match his explicit beliefs.

Later, in his explanation of action, Velleman moves on to narrative understanding, which we have seen is explained through an emotional cadence that ends with a final affective sense of the story as a whole. On the first account we make decisions based on what would make sense to do, given what we represent ourselves as being, but on the latter we make decisions based on what would make sense given our self-narrative up to now.

This narrative account, for Velleman, allows us to make sense of certain human behaviour that cannot be made sense of by understanding agents as *homo*

[47] For discussion of his causal-psychological theory, see Velleman's paper *The Centered Self* (In Self to Self, 2006). For discussion on the contribution of our desire for consistency, see *What Happens when Someone Acts* (1992) and *From Self Psychology to Moral Philosophy* (2000).

economicus[48]. For instance, we try to make the best of our misfortune, make lemonade out of lemons, but, "if life hands you a lemon, the instrumentally rational course may be to throw it away and look for a kumquat instead" (2007, p. 286). If agents are narrative creatures, however, then their life consists of rich emotional meaning, where bad events are made sense of in virtue of what comes afterwards. *This* is why we try to make lemonade, rather than finding a sweeter fruit.

This contrast distils the difference between the account I favour and G&K's account. While I will argue that narrative agency is of central importance, G&K make causal-psychological agency at least as important.

However, implicit in their account, and to be argued for by me, is an idea that these two senses of agency are continuous. Velleman, on the other hand, thinks that a causal-psychological and a narrative account of agency both provide independent contributions to our capacity for agency:

> *To the question how I reconcile narrative and causal-psychological self-understanding, then, the answer is that, to some extent, I don't… At the theoretical level they constitute, in my view, independent and potentially competing modes of practical reasoning*[49] (2007, p. 287.)

On the other hand, if, as I will argue, it turns out that our capacity for causal-psychological reasoning depends on narrative understanding, then whatever contribution our psychological-causal reasoning makes to agency will partly depend on what narrative understanding contributes. This makes

---

[48] *Homo economicus* being a term describing the view that humans are rational, and that their rationality consists in economic reasoning i.e. balancing costs and benefits.

[49] Velleman doesn't make clear why he thinks these types of understanding are independent and sometimes competing. He does, at one point, contrast instrumental reasoning, where it appears irrational to invest in projects that appear to be failing, to a narrative approach. When he does this, he seems to understand instrumental reasoning as understanding "our actions in terms of their motivating aims" (p. 285). Since this is a case of metarepresentation, it could be that he thinks of causal-psychological reasoning and instrumental reasoning as integrated or the same process. In which case, we can understand his statement that narrative understanding and causal-psychological understanding are independent and sometimes competing as supported by the suggestion that causal-psychological understanding and narrative understanding offer competing answers to questions such as whether we should continue to invest in currently failing projects. But it seems false that their answers *must* be competing: causal-psychological understanding can incorporate our awareness of our narrative tendencies. For example, if I explicitly realise that I need to act in ways where my current actions make sense in terms of my past commitments then my causal-psychological understanding does not compete with narrative understanding.

narrative understanding the more fundamental ingredient in our theory of agency, and makes our theory of agency coherent.

2.2. Reconciling narrative & causal-psychological understanding

What may already have started to come into view in the preceding chapter is how these two capacities, that Velleman (2007) thinks of as "independent and potentially competing modes of practical reasoning", are actually integrated. There I suggested that we can understand self-narrative in three ways. First, all narrative understanding, including narratives about lives and events that are ostensibly nothing to do with you, are self-narratives, because all narrative understanding includes a prereflective sense of self. Then there is the type of narrative understanding involved in MTT. That is, understanding narratives about your own past and future. Finally, there are metarepresentational narratives, where you explicitly represent yourself as a psychological being leading your life[50].

Causal-psychological understanding, in this way, is the third type of narrative. We understand ourselves as psychological beings insofar as we understand ourselves as having a particular perspective that is continuous through time. Just like we can wonder about the intentions of characters in novels and soap operas, we can wonder 'what am I up to?'.

To take a reflective stance towards ourselves in this way first requires us to have a perspective to reflect on. Specifically, to have a perspective we identify as *our* perspective. And to have a perspective we identify as our perspective we need to be able to experience our perspective as continuous through time. This is what happens in MTT: we experience our self as the agent/character for whom there are beliefs, motivations, and specific attitudes. We thus implicitly experience ourself as a diachronic agent.

---

[50] There is a further issue of whether even the implicit self-awareness present in all narrative is in some sense metarepresentational, i.e. a representation of a psychological subject. Someone like Damasio (1999) might have such a view where we become prereflectively aware of ourselves only when our brain represents our body as being altered by an object. However, I take this debate, about neural representation, to be orthogonal to my purposes here. For my purposes, metarepresentation means being explicitly aware of representational states, like beliefs, as representational states.

Our ability to reflect on ourselves as causal-psychological beings may involve additional capacities aside from narrative understanding. But without first experiencing the perspective of a diachronic psychological being this stance would be incomprehensible, for there would be no content to our understanding of what a psychological thing is. What a psychological thing has to fundamentally consist in is a thing for which there is a single perspective across time and situations. It is hard to understand something as having a mind that does not involve the idea of an entity that exists through time, makes sense of the world in a relatively systematic way[51], and where the sense-making, because it is that of a single entity, persists through different contexts and times, with the entity. Narrative understanding, where we experience the perspective on a diachronic psychological creature, gives content to our understanding of ourselves as psychological creatures.

We saw this earlier with Samina. Samina, remember, has amnesia but has not lost her ability to understand events narratively. Once she has narrative understanding, the capacity to have a perspective on and in a sequence of events, she knows what it is to exist as a diachronic psychological agent. She doesn't even have to engage in MTT to do this, she could read Harry Potter instead. But it is the capacity for narrative understanding that means that she can represent a being as a psychological being in the first place. She could not represent this, in that 'psychological being' could not refer for her, if she did not first have an implicit understanding of there being things which could have a perspective, and could have a perspective through time. Because she has narrative understanding, that is, because she can experience a perspective as existing through time, Samina can represent that perspective as a perspective. The point here is that the implicit understanding, or the experience of, a diachronic perspective is prior to metarepresentation. Metarepresentations depends on narrative understanding.

We may wonder why it is necessary that Samina has narrative understanding. Why is it important that she inhabit counterfactuals, when she

[51] We can even wonder whether the idea of something making sense implies systematicity. It is hard to understand how we can understand anything as meaningful without some ability to understand how different things are related to each other.

inhabits a sequence of factual situations? The answer is that it would be possible, without the capacity for narrative understanding, that a creature can go through a sequence of events without experiencing diachronicity. What is sometimes called a 'minimal' or 'core' sense of self can consist in an awareness of the moment and perhaps the proximal past and future (e.g .Damasio, 1999; Metzinger, 2004), without any real experience of a continued experience through time. Narrative understanding, by contrast, requires the possibility of a perspective continuing through a sequence of events, events that are non-actual *and* distal[52]. This is the type of perspective we associate with complex psychological beings such as ourselves, beings with desires, beliefs, hopes and goals.

A possible objection to the argument that the subjective experience of a perspective is needed for one to be able to metarepresent, is that it doesn't generally seem necessary to understand what it is like to be something to understand that thing. One can understand a square without inhabiting squareness. Note however, that subjectivity is (in)famously not analogous in this regard[53]. There is something it is like to inhabit a perspective that isn't exhausted by explaining it from a third-person perspective (e.g. Jackson, 1982; Nagel, 1974; Taylor, 1983). Similarly, to theorise about the goals one pursues over time, one must first experience oneself as existing as an embodied perspective through time. Because this is the type of thing it takes to exist for goals to exist. Goals are not free-floating. They belong to some thing for which there is a future. And, as we have seen in chapter 2, to be a thing for which there is a future is an important aspect of diachronic agency.

What I have tried to show above is that we have conceptual reasons for thinking that causal-psychological thinking piggybacks on perspectival narrative thinking. In the strongest form, this is proposed as an argument from necessity: it just doesn't seem possible to conceive of a diachronic perspective if we haven't first experienced ourselves as being a diachronic perspective. At the least, engaging in narrative understanding is a good way to explain how we metarepresent, even if it may not be a necessary move.

---

[52] That is, not just about the happen the very next moment.
[53] To be clear, I think it is possible to acknowledge this while being a materialist.

This is important in the context of what it is to be an agent. We saw in the last chapter that G&K, and those they take themselves as following – Velleman, Korsgaard and Bratman – all argue that our causal-psychological capacities can explain agency. What we have seen now is that we have conceptual reasons to think that such explicit awareness is enabled by narrative understanding. Thus narrative understanding is explanatory how metarepresentation is possible. As we saw in the last chapter, there is some empirical reasons to think that problems with narrative thinking co-occurs with problems with making and acting on decisions. What swings the debate in my favour is that the upcoming arguments that narrative understanding is sufficient for unifying our selves, and that emphasis on metarepresentational capacities doesn't categorise the right people as agents.

There is an issue I have not addressed yet, that I will take up in the next subsection. That is, metarepresentation, through allowing us to focus on our psychological processes, is often taken as necessary for us to be self-consistent, which in turn is needed for agency. It is not (yet) clear how narrative understanding, which does not involve reasoning about ourselves as psychological creatures, can do this.

Yet a final potential flaw in my proposal arises when I combine my reasoning above with a claim from my last chapter, that the ventromedial prefrontal cortex (vmPFC) is important for this kind of narrative understanding, and with the empirical evidence that EVR, who had a vmPFC lesion, performs at least as well as controls on tasks to do with social cognition. Saver & Damasio (1991) conclude that "damage to the ventromedial frontal sector does not compromise the activation and manipulation of neural records that define the varied parameters of social knowledge e.g. social rules, current contingencies, possible response options and future outcomes" (p. 1245). If vmPFC is, on my theory, crucial for narrative understanding, and narrative understanding is a condition on the possibility for metarepresentational narrative, then the capacity of a person who has a vmPFC lesion to engage in reasoning about what people do in what contexts, and what the outcomes are of actions, may pose a problem for

my account. One explanation for EVR's success is that he is representing people's mental state while having a deficit in narrative understanding.

However, because this interpretation is counter to a wealth of other empirical considerations, which are supported by a coherent theory, there is a motive to account for this data through other means. These include:

- Noting that it is not clear precisely what part the vmPFC is playing in humans, and while it seems important in the normal case, it may be that there are other brain regions that perform similar enough functions. For example, Ferstl et al. (2008), found in a meta-analysis that one section of the vmPFC and one section of the dorsomedial PFC (another part of the medial PFC) were active in discourse understanding. Thus it is possible that an adjacent region is also significantly contributing to overlapping abilities.
- It is not clear that EVR is using metarepresentation. The vmPFC does normally seem to be involved in metarepresentation because it is active when understanding another's cognitive perspective (Walter, 2012). In light of this we might understand EVR as predicting other people's behaviour by learning regularities between cause and affect that do not rely on a representation of psychology.
- The evidence above implies, at the least, a developmental story. Here, one first has to understand having a perspective through time before one can understand a metarepresentational narrative. Since EVR's lesion happened later in life, he had developed metarepresentation already, and this capacity might thus be distributed far wider than just the vmPFC. If we accept that certain capacities may be enabled by widely distributed brain regions, then, while a particular area might be initially pivotal in the development of a capacity, and, in the normal case, may remain in use, we can understand how 'graceful degradation' could be possible when one part of the distributed network enabling a capacity is removed.

- Remembering that the brain is highly malleable, it is plausible that some parts of the brain take over activities normally associated with other parts in the case of lesions (Hurley & Noë, 2003).

These explanations are not mutually exclusive. Regions other than the vmPFC are likely involved in enabling the understanding of behavioural sequences, and either through development and/or through a response to lesions, other areas may compensate for the function initially strongly dependent on the vmPFC. While these regions may enable some sense of a diachronic perspective, this may be an impoverished sense, but one that is ameliorated for EVR by a strong behaviourist understanding of human interaction.

Moreover, we have seen in chapter 3 that EVR, in regards to his own situation, can theorise about what to do while not be able to turn that into action. This indicates that while there are surface similarities between his capacity to explain a sequence of events and those without vmPFC lesions, there are some differences too. For most adults, providing a description of how to make friends involves some understanding of how to execute the actions involved in this situation, whereas for EVR it means nothing to him in terms of behaviour. This, I believe, is because of a deficit in how his explanation of a sequence of events is connected to a perspectival, embodied stance, or, at least, how richly embodied this stance is.

Not only is there something parsimonious in explaining why apparently different processes are actually continuous, there is also some additional explanatory value that comes from explaining one process partly in terms of the other. For instance, we can explain how causal-psychological processes motivate: for the same reason that our narrative understanding motivates. This is something that G&K indicate. Causal-psychological reasoning motivates when the decisions we come to through it incorporate a sense that the perspective in the future, for which these decisions will matter, is *ours*. That the actions we take now will, through narrative, affective, embodied continuity, become the future we aim to live.

This is not to deny that there might be something special about psychological narratives. While I do not want to go into depth about this, we get an

indication of what this might be with the example of Pablo who believes he is not sexist but has come to realise he has some implicit sexist attitudes, which in turn cause some sexist behaviours. Pablo's introspective narrative enables him to recognise the distance between his judgements between how he ought to act (not being sexist) and his current behaviour. He can determine ways for his behaviour to be more in line with his narrative, so he doesn't act in sexist ways. In some cases, bringing our behaviour more in line with how we understand ourselves to be may be possible through reflection and self-control. Pablo might think more carefully about whether some of his beliefs and values contradict his egalitarianism in an attempt to be more coherent and renew his commitment to egalitarianism (a strong commitment to egalitarianism has been shown to affect implicit attitudes affecting behaviour (Moskowitz & Li, 2011)). In other cases we may avoid certain environments (e.g. if one is a recovering alcoholic who has realised that when they go to a pub they desire drink) or change our environment (Pablo might plaster counter-stereotypical images of women around his house, as this has been shown to alter implicit gender biases (Blair, 2002)). That is, narratives about our own psychology may enhance our capacity to do what would make sense, given our understanding of who we are.

2.3. Narrative understanding in agency

What I have argued above is that narrative understanding and metarepresentational self-understanding are not entirely distinct. Instead, the latter depends upon, and makes use of, the former. Further, there is some explanatory value once we see it this way: we can understand why our metarepresentational narratives motivate action. Now I explain why narrative understanding, the more basic of the two, is sufficient for agency. That is, the former section established that metarepresentation is enabled by narrative understanding, but it neither established that metarepresentation wasn't a vital development in our narrative capacities, nor explained how narrative understanding could be sufficient for agency.

Remember that, here, agency is doing what would make sense in light of the type of thing we are. Velleman argues that we can do this either through

narrative understanding or causal-psychological understanding. Both of these are meant to enable us to form a coherent perspective from which to experience the world, so that we can choose plans and actions that are consistent with who we are. It is our drive for our actions to make sense that explains how we can act for reasons. We care about being coherent, so it is in light of some narrative understanding of ourselves, or some explicit expression of who we are, that we make plans and goals. For Velleman, these are two separate ways that we can become coherent, although he suggests it is useful to use both. In contrast, what I am suggesting is that causal-psychological understanding becomes possible through narrative understanding.

Similar to Velleman, G&K see agency as involving both causal-psychological and narrative understanding. There are some reasons why one might think that metarepresentation, and not just non-metarepresentational narrative, may be necessary for this. To act in light of who we are requires that who we are is relatively consistent. As Korsgaard (1989) notes, we need to be relatively unified so that we can choose how to act. Otherwise, contradictory plans and projects present themselves to us, and there is no way to resolve them. The fact that we act at all shows that there was a way to decide between alternatives. The question is: what is it that helps us to decide to pursue one course of action rather than all the other possibilities? Explicit understanding of ourselves might avail us here. Because it seems that it is through thinking about who we are explicitly that we are most likely to spot contradictions in our understanding of ourselves and form a coherent sense of self.

In this section, I argue that psychological-causal understanding is not crucial to agency because it is not crucial to having a unified self. Narrative understanding can do the heavy lifting in explaining how we can come to a consistent self-understanding on which we can act on. In this way, I depart from the causal-psychological account that G&K give of agency.

My central observation here is that we don't need to think about ourselves as psychological creatures to have a unified perspective on the world. We can think about the world, and make our understanding of the world consistent. What

follows is an explanation of why narrative understanding of the world makes possible a unified agent.

The first thing to point out, is that the pre-reflective nature of narrative understanding draws self-understanding and world-understanding together. They are two sides of the same coin. This is due to the affective, relational, nature of embodiment: inhabiting a series of events means understanding what those events mean to you. The final perspective we end up with in a narrative implicitly tells us something about ourselves (what something means for us) and something about nature of the situation (the type of stuff out there, where 'type' depends on how it relates to oneself). So self-understanding and world understanding happen in tandem in narrative understanding.

The second point to draw out is that narrative understanding, on my account, involves a recombinant system, and not just affect. I will briefly explain this here, but it will be developed in the next chapter. A recombinant system, in this context, is any system that allows the offline recombination of parts, which may include intentions and perceptions, and means and ends (Hurley, 2003; 2008). Narrative understanding involves employing a recombinant system to take us through the sequence of events. Further, some narrative understanding involves recombinant systems, such as language, that allow us to reflect and draw inferences. In this case, our narrative understanding is under intentional control. This is how it is possible for us to be the kind of interpretive creatures that I claim we are in the next chapter: that the stories we understand, by their nature, are always open to revision. But if this is the case then we need to be clear that 'narrative understanding', when it uses a system like language, is really shorthand for 'narrative understanding and production'. Understanding narratives relies on the capacities needed to produce a narrative. A creature who understands narratives is understanding how parts are organised to produce a whole. Such an understanding can be actively deployed to produce a narrative too.

This is supported by neurological literature. Mar (2004) reviewed and compared stories showing which brain regions are active in story comprehension and production. He found there was large overlap: "few regions appear to be

associated uniquely with comprehension or production" (p. 1424). Further, he claims that,

> *There is a good theoretic reason why narrative comprehension and production should be related…at the level of narrative the ability to organise the meaning of connected sentences in order to form holistic representations for either understanding or communication seems to be a shared necessity.* (p. 1424-1425.)

So, in fact, narrative understanding, to some extent, depends on the capacity for production. To understand a narrative one has to be able to construct for oneself the sequence of events. While this may be done passively, in the sense that one is constructing for oneself a sequence of events through listening to someone else tell a story, it can also be done actively, as when one creates a new narrative.

It matters that we can produce narratives using language, because language – in allowing inferential promiscuity, productivity and systematicity[54] – enables the capacity to create and maintain coherence. However, this is the case regardless of what we are using language to describe. Because we can manipulate language and draw inferences from it, when we use it to understand the world, we come to coherent and novel descriptions of the world. For example, if I believe that all living things die, and I believe that my spider plant is a living thing, then I can infer that my spider plant will die.

Stories have this structure too. Stories do not just make sense experientially, to some extent, they make sense logically[55]. If Harry is the hero, and Voldemort is a baddie, and heroes fight baddies, then Harry (unless otherwise prevented) will fight Voldemort. If heroes are good, and good conquers bad, then Harry will conquer badness. Narrative understanding enables us to form a

[54] Inferential promiscuity means that each proposition can be combined with myriad other propositions to form novel conclusions. Productivity is the emergence of novel conclusions from those prior ones. Systematicity refers to the fact that these inferences are logical relations where, for example, contradictions are looked for, spotted and removed.

[55] It may be objected that the 'twist in the tale' is evidence against this. First, I am not suggesting that narrative is always fully logical. More importantly, an unexpected twist works because we are led down one inferential path, and then we are shown that another one was possible. Events still hang together in the unexpected narrative.

coherent perspective on the world to the extent that it produces an understanding of how events, actors and contexts hang together and causally interact.

It is relevant that, despite the fantasy often involved in narrative, the sequence of cause and effect in a narrative has something to say about how we understand the actual world. Once we have some understanding of how heroes act, and we can identify people in our life with a similar character, we can form certain expectations of how they will act, and what the pattern of cause and consequence will be. For example, if we start understanding our colleague as heroic, and then we get a new tyrannical boss, we might form the expectation that our colleague will stand up for our rights. Or, if I am a freedom fighter, fighting world evil, then I expect to win. If my opponents are the owners of the means of production/terrorists/western colonialists, and they are the source of evil, then I expect them to be conquered.

We are now in a position to see why narrative understanding allows us to create and sustain self-coherence:

i) In narrative understanding, all experiential understanding of the world also involves a pre-reflective understanding of self;

ii) Narrative understanding comes with the capacity to create and maintain coherence in our understanding of the world;

iii) Therefore, narrative understanding enables the creation and maintenance of a coherent experiential understanding of self and world.

Crucially, iii) follows regardless of whether the narratives involve our own lives (MTT) or whether metarepresentation is involved. When and if narrative understanding involves the intertwining of language and affect, even a story that only explicitly focuses on the world, and not the self, will implicitly contribute to the coherence of self. We form a systematic and relatively unified set of beliefs about the way that the world is, and will unfold, through narrative. This has consequences for our own actions. For example, Leïla does not have to believe that she believes that sexism is wrong, for her to believe that sexism is wrong. And if she believes that sexism is wrong, and that Jack is sexist, then she might decide that she should challenge Jack's behaviour. The action that she takes, or is disposed

to take, may well be informed by background narrative scripts about how events should unfold, such as implicitly taking the perspective of a heroine that fights against bad. Similarly, her understanding of what 'sexism' consists of, and why it is bad, and how to respond to it, may involve patterns of cause and consequence that take a narrative form.

What matters about this being narrative is that the process that tells us what events mean for us prompts us to act. Narrative understanding of a scene before us and how it should play out is affective and so perspectival. In coming to understanding how events cohere we also come to a coherent perspective from which to understand, and respond to, those events. We unify our understanding of the world to unify our selves.

What MTT and metarepresentation are meant to enable is a coherent self, the organisation of our ideas and behaviours into a relatively unified whole. But we can see that narrative understanding can also do this. It allows us to organise our perspective and motivates us to act on the understandings we come to. On a narrative using metarepresentation, Pablo believes that he believes that sexism is wrong, becomes aware of his implicit sexist attitude and takes action accordingly, according to his narrative understanding of how anti-sexist people behave. Or behaviour can be consistent with beliefs and attitudes without metarepresentation, but through the systematicity and productivity of language, combined with the perspectival nature of emotion. As when Leïla questions Jack on his beliefs and behaviour because he acts in a sexist way.

One may worry that metarepresentation does increase our ability to be consistent through time, because someone who metarepresents thinks of themselves as one being. Without this, one could have many, sometimes conflicting, context-sensitive narratives. The potential problem is that this might make one less likely to act consistently through time.

What we have to be clear on here is the possible ways of accounting for agency. Velleman's basic notion is that an agent is a creature that wants to be intelligible (to itself). At its most basic, this does not have to involve diachronic agency in the sense of making and following long-term plans – what it does require is a way to have a coherent set of values and attitudes through which some actions,

and not others, make sense. This will mean that a creature acts relatively consistently through time to the extent that those values and attitudes persist. For diachronic agency in the sense of pursuing plans and goals, some explicit recognition of self seems necessary so that one can identify ones actions as part of a wider plan through time. However, this understanding of self may not be metarepresentation. It can be the identification of a body as our body, along with the roles and responsibilities that we have.

However, it seems that metarepresentation will sometimes help us act consistently through time. If we are away that our current behaviour is being influenced by factors that distort our perspective, for instance if we are drunk or furious, metarepresentation is a useful way to understand that our current perspective is uncharacteristic of our general point of view. This may often help to steer our behaviour in a way that is more consistent with our norm. Recognising that the red haze in front of my eyes is a result of rage because the toddler I look after has been obstreperous all day – and that I should probably sit down and recollect myself before I pursue my interactions with him – is a good way to prevent myself transgressing from my general understanding of good childcare. An understanding of my own psychology is very helpful here and helps me to be consistent. Recognising times when metarepresentational capacities expand our agency, however, does not mean that metarepresentational capacities are a minimal condition of agency.

Yet metarepresentation may be limited in the extent that it facilitates consistency. If it were not limited, then understanding ourselves as compassionate, rational, and independently-minded agents would be enough to prevent us acting in the way that we tend to in the Milgram experiments[56] (1963 & 1974), and that doesn't seem to be the case. It seems situational factors have quite a strong affect on us, and therefore contribute to us acting inconsistently through time, regardless of metarepresentation. Furthermore, if we can compare and contrast narratives

[56] These are the famous experiments where someone who appeared to be a scientist informed participants that they should press a button if another person, who was being asked questions, got those questions wrong. The participants were told that this button would cause an electric shock, although this was not true. Most participants did what the scientist had asked them to do.

about the world, we can pick out inconsistencies in our understanding of the world. Because the idea is that it is the consistency of our perspective that drives the consistency of our actions, such a world-directed perspective should contribute to us acting consistently through time.

I think there are further merits to understanding narrative understanding, rather than metarepresentation, as fundamental to agency, as we shall see next. It allows us to respond to Galen Strawson's concern that narrative understanding is not an important part of most people's psychology, and it does better at explaining our intuitions concerning who counts as agents. This is what I turn to next. I start of with elaborating on Strawson's central insight, before explaining why a more precise theory of narrative understanding defuses his criticisms of its contribution to agency. This allows me to further illustrate my view, and respond to a prominent objection to narrative accounts of self.

3. Criticism of narrative agency

3.1. Counting in the right individuals

Galen Strawson, in his 2004 paper 'Against Narrativity', argues that explicit narrative understanding of oneself is not important for selfhood. He claims this must be true because some people have coherent selves without engaging in (much) self-narrative. In particular, they don't engage in any non-trivial sense of narrative understanding. A trivial sense of narrative is claiming that, "making coffee is a narrative that involves Narrativity, because you have to think ahead, do things in the right order, and so on, and that everyday life involves many such narratives" (p. 430). For him, "Narrativity", in the non-trivial sense, is the ability, consciously or not, to seek coherence and form in the recounting of the events of one's life, "One must have some sort of relatively large-scale coherence-seeking, unity-seeking, pattern-seeking, or most generally [F] *form-finding* tendency when it comes to one's life, or relatively large-scale parts of one's life" (p. 441). One must seek that form finding through a story-telling structure, detecting "developmental

coherencies" (p. 443) and "deep personal constancies" (ibid.) that are present through one's life.

Strawson objects to Narrativity being necessary for selfhood on the grounds that it isn't needed for self-understanding, and is quite often a hindrance to it. Understanding oneself narratively can obscure self-understanding due to the kind of "changes, smoothings, enhancements, shifts away from the facts" (p. 447) that occur when we try to make our life into a narrative. So this activity, of weaving life into narrative, often obscures our nature rather than illuminating it. For Strawson, self-understanding need not involve the putting together of bits of one's life into a narrative. Even in psychotherapy, one can benefit from learning about how effects in one's past are related to one's current anxieties without having to understand one's history narratively[57]. Form-finding, if necessary for agency because it is required for self-understanding, does not have to occur in the form of Narrativity.

Strawson's comments about human psychology are important for this project, because if narrative understanding is not a common and/or useful feature of our thinking and decision-making, then it is hard to claim it plays an important role in agency. Additionally, his claim that Narrativity obscures self-understanding may appear to contradict my claim that narrative understanding is *the* central way that we have (implicit) self-understanding.

I will later go on to show why Strawson's claims against the contribution of Narrativity to our psychology can be defused with a more detailed and determinate understanding of narrative. For now I want to focus on what is right about Strawson's analysis, which has implications for G&K's use of metarepresentation in their theory of agency.

Strawson's argument denies that metarepresentational narratives are necessary for selfhood. He claims that trying to make sense of life as a whole is not something he does, nor something that many other people do. But failing to do this does not stop a person from being an agent. A psychological activity that many adults never engage in seems like a bad candidate for an activity crucial for agency. What I want to point out now is that it is only when we exclude

[57] Strawson doesn't elaborate on how this is possible and why this possibility differs from a narrative approach to therapy.

metarepresentational narrative from a minimal account of agency that we delineate the right creatures as agents.

If we take Strawson at his word, and agree that many people don't normally engage in metarepresentational narratives, it seems wrong, on that basis, to not consider them agents. They can consider counterfactuals and decide how to act on that bases. They can have a relatively unified outlook with which to assess and respond to events. So they can have a sense of themselves and what matters to them, and, through this, act based on these attitudes.

Moreover, we have more evidence than Strawson's testimony that metarepresentational accounts can't make the right delineations between agent and non-agent. The anthropologist Maurice Bloch (2011) has claimed that there are societies where people only occasionally engage in metarepresentational thinking. Bloch notes that anthropological data has exposed that in some societies, people are generally understood as "points in social systems, while their internal states, their intentions, their absolute individuality and personal desires are irrelevant" (p. 4). This is a problem for a metarepresentational account, because it implies that there are many adult people who are severely limited in the extent of their agency, not because of any obvious impairment, but because they rarely explicitly understand themselves or others in psychological terms. Nonetheless, Bloch supports the notion that in all societies, people have narrative selves that go beyond a mere experience of themselves as existing through time, to include an understanding of themselves as actors within a social group, with certain roles and responsibilities. Such narrative understanding needn't involve explicit awareness of psychological states.

Bloch's levels of narrative understanding involve a different taxonomy from the one that I am proposing. However, what is provided through Bloch's account is empirical data that confirms Strawson's claim that many people rarely engage in metarepresentational narratives. Yet these are generally people we would count as agents and who have narrative understanding. So a narrative understanding account, but not a metarepresentational account, counts the same creatures as agents as an everyday understanding of agency. This is not to say that our everyday understanding of agency should be dogmatically defended. However, if

two accounts both explain some everyday category, but only one also ends up counting in the same individuals as tokens of that category as our everyday understanding, then we should prefer the one that meets both conditions.

Further, Bloch's analysis gives us a useful way of understanding how we can act on goals and plans without metarepresentation. If we have an implicit understanding of our selves as existing through time, and an explicit understanding of our roles and responsibilities, then we have a way to plan without metarepresentation. For example, we might remember "I told my friend I would make them dinner tonight" and that understanding of our plans gives us a way to act on previous decisions.

However, one may counter that it is not the case that people who generally understand themselves narratively through an implicit understanding of themselves existing through time, and through seeing themselves as actors with certain roles and responsibilities, can't metarepresent, just that they generally don't. Both Bloch and Strawson admit it is a matter of degree. So, one could say, what is important for agency is that one has the capacity, not that one exercises it.

This ignores the fact that metarepresentation is given a functional role within G&K's account of agency. G&K claim that what MTT contributes to agency largely consists in endowing a creature with self-knowledge in the light of which the creature can act. Their argument is that through thinking about our autobiography we make sense of ourselves, and can choose actions that are in accordance with our self-conception. My argument is that we should resist giving metarepresentation in MTT and narrative understanding such vital importance. It is clear that those creatures we count as agents frequently engage in narrative understanding but not necessarily metarepresentation. And narrative understanding, as we have seen above, allows us to have a coherent self-understanding which we are motivated to act in line with. Metarepresentation, although it enables new ways of forming coherent self-understanding, isn't necessary for this.

3.2. Defending narrative understanding

Once we have the elaborated description of 'narrative understanding' I have presented, and we concede Strawson's point about metarepresentational narratives, we see that his objections towards narrative agency dissipate.

Strawson's confusions about what defines 'narrative' are shared by those he engages with. We have seen above that Velleman alternated between causal-psychological and narrative explanations until Mackenzie (2007) prompted his clarification. Similarly, G&K talk about diachronic experience and autobiographical narratives in parallel without examining how they are related.

So we need to have a clear account of narrative understanding before we can assess which of Strawson's criticisms are valid. We now know that narrative understanding can be 'self-narrative' minimally: a self-narrative in that it contains some sense of self and what things mean to you. We can give a more concrete descriptions of what narrative understanding consists in, other than Strawson's assertion that it is 'form-finding'. While 'form-finding' in the sense of inferential coherence might be one characteristic of narrative understanding, what is intrinsic to narrative understanding, and especially interesting about it, is emotional cadence. If by 'form-finding' Strawson means explicit consideration of one's life and psychology, this isn't necessary for narrative understanding. What becomes apparent, then, is that both sides of the debate have suffered through a lack of clarity of what defines 'narrative'.

Once we have this definition in place, we have the tools to respond to Strawson's critique. First, Strawson's worry about non-metareprentational narratives being trivial dissipates once we have a more determinate sense of what we mean by 'narrative understanding'. Now we understand that what we mean by narrative understanding involves emotional cadence and a recombinable system. On this account, coffee-making, or thinking about coffee-making, is narratively experienced to the degree that it is an affective endeavour understood through using a recombinable system. So, if one quite likes coffee then these activities may be experienced only slightly narratively, but if one is addicted to coffee or hates it, or is a passionate coffee connoisseur, then it will be experienced more narratively.

Further, as mentioned in the last chapter, although an emotional cadence is characteristic of narrative understanding, once that exists, there are other dimensions to assess the extent to which a sequence is narrative:

- The extent to which our coffee narrative results in holistic understanding. The more parts of the narrative that are consistent with one another, the more narrative it is.
- Whether this holistic understanding involves complex parts coming together. If you are a coffee connoisseur, and there are many parts and variables to your coffee making, then your narrative may be richer gestalt.
- Whether we understand this narrative through the type of recombinable system that allows us to reflect, i.e. language. If this is case, then it is the type of narrative understanding specific, as far as we know, to human agency. More will be said about this in the next chapter.

Interestingly, once this is explicit, I am able to explain what is troubling about the coffee-making example. It is perhaps a combination of a lack of obvious affective significance, and a rather simple gestalt, that makes the coffee-making narrative seem deficient: without the context of the rest of our day, and how coffee fits into it, a coffee-making narrative is generally rather short, simple and dull. As soon as we link it to our prior drowsiness, and our future invigoration, we get both a more obvious emotional cadence, and the coming together of complex parts to form a whole. That I am willing to call even a rather decontextualized and trivial understanding of coffee-making minimally narrative is due to commitments that I explain in chapter 6 concerning how our understanding of sensorimotor activities is interdependent with affect.

Given that this theory of narrative understanding can explain why we find some sequences of events to be more obviously narrative than others, it does not make narrative understanding trivial, it just makes it relatively ubiquitous. The charge of triviality would stick only if our explanation of narrative understanding had no interesting features. Compare this to a philosopher who thinks all cognition

is conceptual. We would only call this theory of cognition trivial if such a philosopher could not explain what characterises 'concepts'.

Further, with the distinction between different ways narrative understanding can contribute to self-understanding, we saw above that we have the capacity to explain Strawson's intuition that explicitly thinking about one's whole life is not the only way we come to self-understanding. His resistance to narrative understanding is dependent on a notion that the only non-trivial sense of narrative understanding is one where what one is trying to understand is one's life as a whole. Particularly, he is concerned about the idea that introspecting on one's own motives and beliefs is necessary for a sense of self. That is, he has a problem with the necessity of metarepresentation, and I've upheld that objection while maintaining that narrative understanding is still important for our sense of self.

It is also questionable whether non-narrative self-understanding does exist. One might think that narrative understanding seems limited because our ability to imagine what we would feel in the future may be markedly different from how we would feel in a certain situation. For example, I might imagine that if a scientist were to ask me to electrocute someone for answering a question wrong (see the Milgram experiments, 1963 & 1974), I would feel horrified and therefore refuse to do it. My knowledge of the way the majority of people act might allow me to make a different judgment about my future self, and to imagine different possibilities. Strawson might point out that, in these cases, my theoretical understanding of myself may be more accurate than my narrative self-understanding.

While I agree that friends, science and sociology are important sources of information about one's self, I don't think that that the contributions these sources make to our self-understanding are non-narrative. When thinking through the idea that myself or another person would electrocute someone for purportedly getting an answer to a question wrong, my experience of revulsion expresses my understanding that this is an unacceptable way for someone to act. There is an emotional cadence experienced when thinking through this sequence of events, and that is why we have a sense of what these events mean for us.

When I apply Milgram's finding to myself it helps me understand that, as a human, someone's status will influence me regardless of whether I want it to or

not. It is the embodied understanding of events that allows me both to judge what I think of those actions, and to take the perspective of the participant and the person that could have ended up being electrocuted, and that enables me to apply this factual understanding to my own possible behaviours and their consequences. I have widen my repertoire of what it is I can imagine myself doing, in a way I couldn't if such imaginings were purely intuitive. Nonetheless, these imaginings are affective, and therefore understood through narrative.

Therefore, what Milgram's psychological study adds to my self narrative is a belief that my behaviour is not only caused by my explicit values, but also by more implicit factors, such as whether I recognise another as more authoritative than myself. So narrative understanding can lead to self-understanding through incorporating factual or theoretical knowledge that contradicts my pre-theoretical understanding of myself. What scenarios like our understanding of Milgram experiments show is not that narrative understanding is limited, but that our intuitive ideas of ourselves can be.

Note that much of our understanding of the Milgram experiments can be learned though a minimal sense of self-narrative. Thinking through the Milgram experiments involves a certain emotional cadence, but it doesn't necessarily involve explicit reflection on who I am as an agent, or on my life as a whole. I may not use the Milgram experiments to explicitly reconceive my own motives, instead I could incorporate it into a narrative about the world that people generally do what they are told to do if they are told to do it by someone who is an authority, and that some people with authority are dangerous. Here, I add another dimension to my narrative that could lead me to feel fear in response to authority figures, which is part of a skeptical attitude that increases my attention towards their behaviour, alerting me more easily to those authority figures who should not be trusted, and maybe causing me to avoid such people when I identify them. My attention, in this case, is directed at the world, and my sense of self is present pre-reflectively. Similarly, my understanding of the scientist who orders the electrocution and the participants who obey them may include an implicit sense of their perspective. Thinking through the experiment is a type of narrative understanding that affects my future reactions, whether or not I envision the

experiment as part of my life, or use it to explicitly think about what type of being I take myself to be. It does require me to see other people in terms of their roles (including their status) and responsibilities, but it does not require me to explicitly represent them as psychological things.

What surfaces in applying a more precise and detailed account of narrative to Strawson's argument is that there is less need for dispute than initially appears. Strawson's claim about metarepresentation has some weight. His further claims about narrative understanding in general do not follow from his scepticism about metarepresentational narratives. But the confusion is understandable when considering that his interlocutors slide between these senses of narrative at least as much as he does.

4. Conclusion

This chapter shows why my narrative understanding account of agency is preferable to that of G&K's metarepresentational account. It provides a minimal account on how we can act rationally, that is, how we can act in a way coherent with our sense of self. This minimal account has explanatory value: it explains how it is possible for us to take a reflective stance on our own psychologies and why it is that such a stance can motivate us to act. So we can see how narrative understanding is a fundamental capacity that metarepresentation depends upon. Because of this, any contribution that metarepresentation makes to our rationality is partly because it is an elaborated form of self-narrative. And, unlike a metarepresentational account, my account makes the right claims about which people we should count as agents.

The context this chapter falls in, though, is a thesis that is providing an alternative account to Prinz's of the role of emotions in acting for moral reasons. Prinz, remember, thinks that emotions are the constituents of our moral judgements, which count, to Prinz, as reasons for which we act. He thinks that emotions can play this role because they are caused by our deliberative processes and, therefore, emotions are a result of our agency. So, the question now is, why

does my narrative account of agency have anything specifically to with our capacity to act on moral decisions?

Narrative Moral Agency

Being your own story means you can always choose the tone. It also means that you can invent the language to say who you are and how you mean. But then I am a teller of stories… so from my point of view, which is that of a storyteller, I see your life as already artful. Waiting, just waiting, and ready, for you to make it art.
Toni Morrison, Wellesley commencement address, 2004.

1. Introduction

I have proposed that narrative understanding depends upon experiencing events in terms of a certain type of emotional structure. An emotional cadence composed of one emotion leading into another in a partially predictable way is constitutive of narrative understanding. At the same time, the emotional component of narrative understanding is offered as an explanation of how narrative is embodied and perspectival. We inhabit narratives.

In this chapter I am going to relate this to moral agency, and this in turn will have implications for my main interlocutor, Prinz. In particular, I will be presenting an alternative to his proposal that deliberative reasoning is not constitutive of moral judgements, but only an important cause. Instead, I propose that moral judgements are the articulation of narrative-level phenomena, that are jointly constituted through affects and concepts. The stories we tell and understand provide a relatively unified network of attitudes that co-constitute the moral standpoint on which we act. The activity of making these attitudes explicit is the activity of making a judgement.

On this account, our narrative understanding always involves conceptual capacities, and making a judgement requires exercising these capacities. This means that the division is not between deliberative and emotional capacities, but between the passive employment, and active deployment, of affective-conceptual capacities.

So, as a response to Prinz's proposal, my proposal will sit within an alternative framework concerning how we should understand agency, concepts and judgements, and the relationship between these ideas, which will be outlined

shortly. Since my presentation of this alternative does not involve a refutation of Prinz's proposal, this chapter should be understood as presenting a competing and compelling theory, rather than a simple dismissal of Prinz's theory.

Picking up on the threads of chapter 2, as well as the conception of narrative understanding as self-understanding developed in the chapters 3 and 4, I develop the view that a certain type of narrative understanding provides a relatively coherent moral viewpoint through which we make judgements and act. In this chapter we will see that this viewpoint is conceptual in the sense that it is possible – although perhaps not easy – to articulate and reformulate. Through this possibility of articulation and reformulation, we can say that our moral viewpoint justifies our actions, and that we are responsible for our moral viewpoint. Charles Taylor (1985) helps elucidate why this viewpoint is a moral viewpoint. In short, it is because our evaluations of what is of higher value, what is worthy or base to us (i.e. what Taylor calls "strong evaluations"), are formed through a web of affective moral concepts.

There are several reasons for why we should accept this view. First there are philosophical reasons to think that moral judgements are conceptual, as outlined by McDowell. These reasons act, alongside phenomenological considerations, as a motivation for accepting the theory of moral agency proposed by Charles Taylor below. Included in these considerations is the observation that our moral reasoning is continually affective, rather than affective only at isolated moments. Finally, there are empirical considerations that support this theory of moral agency and further incentivise Prinz to resolve his ambiguity towards understanding moral concepts as affective. The empirical considerations are results that show how narratives and moral judgement co-emerge, and that abstract concepts, which I suggest includes moral concepts, are affective. The claim is that there is a fit between what McDowell's theory leads us to expect, the phenomenology of agency, and empirical considerations, if we hold the narrative theory of moral agency. That the narrative theory of moral agency renders all these considerations intelligible is meant to act as a justification for the theory.

As we can see, this chapter both elaborates on why narrative understanding is necessary for moral agency, and explains how a narrative understanding view of

moral agency can accommodate evidence that Prinz's theory isn't obviously compatible with. Prinz, remember, thinks that our deliberative capacities are important causes, but not constituents of, moral emotions. By the end of this chapter I hope to have shown that we have reason to adopt an alternative to this position.

2. Contrasting conceptual frameworks

2.1. Explaining the commitments

To begin this chapter, it is worth getting the main issues of contention on the table. It is important to get these clear, so that we can keep site of the key issues for the rest of the chapter. What I want to do now is recap some of the issues raised in chapter 2, and explain some issues further. Once we have these explained, I will summarise what this means for how my framework differs from Prinz's.

First, lets refresh our memories of Prinz's theory of moral judgements, how these relate to concepts, and the reasons we have to question his account. As we saw in chapter 1, Prinz seems to call emotions conceptual, while denying they are involved in deliberative reasoning. This creates an uneasy hybrid, where emotion both are, and are not, concepts according to Prinz's own theory of what a concept is. In chapter 2, we saw that if we take Prinz seriously as aiming to give us an account of moral judgements, we have reason to think that concepts are constitutively involved. We have an additional reason to give emotions full conceptual status, on this account, because for a process to be capable of being involved in judging, it must be capable of being involved in our deliberative capacities of giving and asking for reasons.

In chapter 1 we saw that one part of Prinz argues that emotions constitute moral concepts. He says this because he wants to explain how emotions can constitute our moral knowledge, and since epistemic states such are judgements are conceptual, and he thinks that moral judgements are constituted through moral emotions, it seems that moral emotions must constitute concepts.

However, we saw that this is in tension with Prinz's sentimentalist leanings, where what separates sentimentalism from rationalism is that making a moral judgement isn't the exercise of our deliberative capacities. This, combined with Prinz's theory that concepts are representations that are consciously manipulable, creates a tension. In his interpretation of dumbfounding results and his description of how emotions are merited, he creates a chasm between deliberative reasoning and emotions, where the former causes the latter. However, if emotions are caused by, rather than participants in, deliberative reasoning, it is hard to see how emotions (according to Prinz's *own* theory of concepts) could constitute moral concepts. Moreover, if emotions were participants in deliberative reasoning, then Prinz would be as much a rationalist as a sentimentalist. It is presumably for these reasons that he equivocates on whether emotions constitute concepts.

We saw in chapter 2 that having only a causal connection between our deliberative activities and our moral judgements can be attributed to a commitment to bald naturalism, a commitment we need not hold. Bald naturalism is a type of naturalism that assumes that we can understand epistemic states in purely causal terms.

To understand how we could be a naturalist and hold something different we need to understand the difference between a constitutive explanation and an enabling explanation. While a constitutive explanation is an explanation of what characterises some thing, an enabling explanation is an explanation of what mechanisms would enable a thing like that to exist. So, what characterises my friend Asher is that he is very funny, and what enables him to be funny may be explained in different types of causal language including, but not limited to, neurological mechanisms.

A bald naturalist understands terms such as 'judgements' as both constituted, and enabled by, causal language. We can resist this, while still being naturalists, by claiming that judgements are constituted by their capacity to justify our beliefs and actions, and can only be explained in causal terms if that explanation is an enabling explanation.

We might think this is true because we have not *yet* been able to understand judgements through causal language. Further, we may suspect there is

a fundamental disconnect in explaining things like judgements, that can be right or wrong, and causal mechanisms, where it is hard to understand how they could be right or wrong. The idea here is that normative language – such as that of judgements, knowledge, responsibility, concepts and reasons – are used to characterise what we mean by agency. Causal language, as I noted in chapter 2, does not seem normative in a way that would allow us to characterise agency.

A similar explanation is offered by Rorty (1980). He argues that to have knowledge that something is the case is a claim about the relationship between a creature and proposition. To assess whether a creature knows that something is the case is to assess whether a creature has a justification for believing a proposition. But "it is rarely the case that we appeal to the proper function of our organism as a *justification*"(p. 141, original emphasis), because an assessment of whether a creature has some knowledge that some proposition is true concerns the logical relations between propositions, rather than a mechanical explanation.

In this light, Prinz runs into additional difficulties, where his idea of emotions as both epistemic states, and caused by, rather than part of, our deliberative agency, would be classed as a mongrel concept.

Prinz's moral judgements are things we can be held accountable for (2007, p. 114). And we can hold a person accountable for a moral judgement because moral judgements are merited, that is they fall under rational deliberative control. The sense in which they fall under our rational control is that they are caused by rational processes, rather than being justified by rational processes. I say emotions, on Prinz's account are caused rather than justified, because according to McDowell, if we were to say that emotions could be justified, we would be claiming that they could be articulated and thus take part in deliberative reasoning. This is not a commitment Prinz appears to have.

Because of this, for McDowell, Prinz's account of emotions being caused by rational processes takes them outside of the activities involved in agency. For McDowell, agency is defined through the capacity to justify, rather than causation.

Remembering that it is conceptual processes that participate in reasoning, Prinz's claim seems to involve the idea that emotions have some conceptual and nonconceptual characteristic. And they have these characteristics both in relation

to his own theory and, as we are focusing on now, in relation to the type of naturalism that is sceptical of causal explanations being a constitutive explanation of agency. In this framework, Prinz's emotions are "mongrels resulting from a crossbreeding of two ideas" (Sellars, 1956, p. 132). They are conceptual, according to Sellars, in the sense that they are appraisals that can correctly or incorrectly represent a situation. This is a feature of emotions for Prinz: "if emotions represent concerns, they can be correct or incorrect. If Jones fears the captive snake, she is literally representing the snake as dangerous. This is an error[58]" (Prinz, 2007, p. 64). However, unlike a conceptual process they are not justified or justifying but caused and causing. This presents a problem for Prinz, because this move is a switch to a causal mode of explanation. But, according to McDowell, to count as judgements, and to really be seen as correctly or incorrectly representing our situation, emotions must fall into the space of reasons. However, when we are understanding a process as falling into the space of reasons, we are no longer applying causal explanation (McDowell, 1994), as explained above.

One might respond that 'merited' for Prinz just means caused in the right way: caused by deliberative reasoning. But the point is that when we understand something as wrong or right based on only including causal explanation we lose something understandable as a judgement. Judgements are an activity that agents engage in, in that they can be questioned, clarified and justified. They are explained through normative language, but this, arguably, takes judgements out of a realm of causal descriptions. In losing this sense of judgement, we also lose the agent, because the behaviour we are holding a creature responsible for is no longer something that a creature has done, but something that happens to a creature. Because, on Prinz's view, the moral emotion on which we act is caused, rather than capable of being part of, deliberative reasoning. Remember, however, that if our behaviour is just something that happens to a creature, then we don't seem to be explaining agency any more. When we lose normative language, and characterise a

[58] Note that Prinz does not think this feature of emotions makes them conceptual. Sellar's argument is that Prinz is wrong in this respect. Because Prinz characterizes them as epistemic states, capable of justifying or failing to justify, they ought to be understood as conceptual. Or, as McDowell would put it, part of the space of reasons.

process as mechanistic, it is unclear why we should say the creature is responsible for the behaviour caused by that process[59].

This argument, however, relies on philosophical commitments that I explained in chapter 2, namely that we should not be bald naturalists, something that Prinz and I appear to differ on. Unlike the bald naturalist, McDowell argues that there is another legitimate type of naturalistic explanation. This type of explanation concerns the understanding a creature as existing within a certain mode of life. In the case of creatures that are agents, this is the mode of life of following goals, articulating beliefs and justifying judgements.

Since I will not offer further justification for this switch of underlying assumption, the perspective of moral judgements I develop in this chapter will not directly disprove Prinz's theory. What has become apparent through the juxtaposition of Prinz's position with McDowell's, however, is that it is not necessary to adopt Prinz's conceptual framework.

On the philosophical picture I am adopting, if we are in the business of explaining moral judgements then it looks like our explanations of what enables us to judge needs to stick to the space of reasons. We need to make judgements the activity of an agent rather than something that happens to an agent. My proposal is that a narrative understanding account of moral agency allows us to see the moral perspective through which we understand the world as formed of affective concepts. It may not be that we always use this perspective to make explicit judgements, but because narrative understanding is a conceptual activity, we can draw on narrative understanding to form explicit judgements and reformulate them.

Now we've got the explanations of two different philosophical frameworks on the table, I want to take an interlude to explain key terms. First, I will summarise how these different frameworks result in the same terms having

[59] It would be a mistake to contrast this view with a compatabilist world-view, since the view I am proposing is coherent with a type of compatabilism that states that a mechanistic enabling explanation is compatible with agency. It is true that this view would not be congruent with a type of compatabilism that claims that we can characterise agency through causal processes. Nonetheless, my starting point was to characterise agency as the ability to act for reasons, something that is widely accepted. This, I am now claiming, is to characterise agency in a non-causal way.

different meaning. This will also anticipate some of the arguments to come. The point is to gather into one spot some quite expansive discussions. Second, I want to introduce some new terms that are relevant to my project.

2.2. Interlude: understanding key terms

There are at least four key terms that need to be defined: agency, judgement, concepts and rationality. These are interrelated in both Prinz's theory and the one I am proposing.

'Agency', in both cases, starts off roughly the same, but is inflected with how it is understood in relation to the other terms. For both theories, an agent is a creature that we hold responsible for their actions. Similarly, there is agreement as to why we hold a creature responsible for their actions: because they are rational. In both theories, the actions for which we hold an agent responsible are those motivated by judgements. In both, judgements are conceptual. Finally, in both theories those judgements are related, in important ways, to an agent's rationality.

Here, however, things start to diverge.

How judgements are related to rationality differs. For Prinz, judgements are "triggered" by our rational capacities. It is enough for judgements to be triggered by our rational capacities for us to say those judgements are merited. For McDowell, and others, judgements are justified by reasons, but not in the sense that they stand in a certain causal relation to those reasons. Instead it is in the sense that we can assess those reasons in deliberation. Once we start using causal language, we have switched out of the language of reasons.

Further, these differences are related to differences in the understanding of 'concept'. For Prinz, a concept is a representation we can intentionally control, and it is because of this that there is a tension in his idea that moral emotions constitute moral concepts. When he explains what moral judgements are, he positions them as caused by processes he believes we can control (i.e. deliberation), rather than part of those processes.

On the view I am adopting, as proposed by McDowell and (as we shall see) shared by Taylor, conceptual activity is any activity that we could, if we wanted to, reflect on and articulate in deliberation. On this account, we do not have to take a

stance on whether we can wilfully manipulate concepts. Instead, whatever it is that we are doing when we are trying to explain or justify our beliefs and actions, is a conceptual activity. Such activity is defined as a certain type of practice, rather than as the manipulation of mental representations. Our experiences are shaped by concepts, and themselves are conceptual, in the sense that they have the potential to be included in the practice of giving and asking for reasons.

Because of this, for McDowell and Taylor, our rationality is defined through our practices as normative creatures, rather than through the manipulation of mental representations. We are rational because we can give reasons, and the process of giving reasons involves actively using our conceptual capacities, that is, making explicit what we believe and why.

'Moral judgements' on my view are activities of a conceptual creature, while for Prinz they are emotions that could be triggered by reasoning. While for Prinz, we can have an unarticulated judgement in the form of an emotional reaction (2006, p. 34), this is not the case on my proposal. For me, while we act for reasons when we act on our narrative understanding, our narrative understanding does not always take the form of a judgement. Such narrative understanding would only be a judgement if, through engaging in the practice of giving reasons, we make it explicit. Our narrative understanding forms the reasons on which we act, because we could formulate our narrative understanding as a judgement if we wanted to.

Finally, not only will I be using the same terms in slightly different ways to Prinz, I will also introduce some new language. Throughout the rest of this thesis, I use the word 'expression' to replace Prinz's use of 'representations', a contrast that Taylor makes concerning how to think of mental processes in general. However, I will be adopting the term 'expression' to refer to emotions in particular. Emotions are expressive rather than representational in the sense that they do not point to something wholly outside themselves. That is, expression is an activity where the act of revealing coincides with what is being revealed. Importantly, emotional activity is always embodied, so that meaning cannot be separated from the matter involved in the activity. This contrasts with 'representation' where meaning arises due to some inner activity that bears a particular relation with some outer object.

As an example of representation, Taylor uses the example of a barometer which "'reveals' rain indirectly" (1985, p. 91). That is, to get meaning from a representation in a barometer we need to make an inference between the sign and what it might be a sign of. In contrast,

> *When I make plain my anger or my joy, in facial or verbal expression, there is no such contrast. This is not a second best, the dropping of clues that enable one to infer. This is what the manifestation of anger or joy* is*.* (Ibid, original emphasis.)

Similarly, we might think that the materiality of art does not just represent some object beyond the medium. Instead of having to make an inference from the painting to what it means, the medium expresses the meaning itself. The paint and canvas in *Guernica* reveals the terror of war and fascism.

I will also be using the words *language* and *recombinant system* in particular ways. What I mean by recombinant system is any process that possesses at least a coarse-grained recombinant structure between intentions and perception, and means and ends, and which can be taken offline to be used in simulations (Hurley, 2003; 2006; 2008). A creature that uses this type of recombinant structure "distinguishes ends from means, recognises that there can be different means to the same end, that the same behaviour can be means to different ends; and that the same behaviour can be an ends or a means" (2006, p. 148). When such a creature can take this process offline to simulate the relationship between their goal, their context, and what means would achieve that goal in that context, this creature is using a recombinant system. A recombinant system need not be comprised of symbols (e.g. like words or numbers), it can consist of the capacity to simulate perceptions, actions and goals.

Language is a type of recombinant system here. It is the type of recombinant system that enables us to articulate a sense further. Language not only enables us to make inferences, but it enables us to see our inferences as inferences. So it is a system where questions of clarifying and justifying are appropriate. It does not obviously have to consist of words (including hand gestures in sign language) as we normally understand them, although, given further argument, we might conclude that the functions of language put strong

limitations on the form it can take[60]. As I understand it, this is similar to what Charles Taylor means by language. I now return to the issue of moral judgements, and why we should see them as constituted by our narrative understanding, which is both conceptual and affective.

3. Moral narratives

3.1. Strong evaluations and narratives

We saw in summary above, and in more detail in chapter 2 that we have philosophical reasons for thinking that moral judgements are conceptual. Perhaps it is in this way that they can be understood to be reasons for action. What is left out of this consideration is a detailed positive theory of how we make moral judgements. Below, I want to show how a theory of embodied narrative understanding can contribute to a theory of moral judgements, and why it is supported by a phenomenological analysis. We get clearer, I suggest, on why the theory of narrative understanding helps explains moral agency, if we follow an analysis by Charles Taylor (1985) of moral agency. He explains why agency integrates our conceptual and affective resources. However, it is worth explaining what ingredients we already have in a theory of narrative understanding that may help explain moral agency.

To see why narrative understanding can be moral, we return first to its affective characteristic, where affect here consists of an embodied appreciation of what some object (a scene, a person, the world in general) means for us. Or, in the language we have just introduced, our affect expresses what matters to us, how we understand an object as bearing on us. Affect therefore comes together with a basic sense of value. Affectivity discloses to a creature what parts of the environment are important to them and how

[60] Conversely, it may be that a recombinant system, of various grains, might allow for clarifying and justifying, in the sense that it might be expressed or altered by a creature if the creature meets normative pressures. 'Language' would turn out to be a misleading word, if a broad-range of capacities quite different from what we understand as language fulfil this role.

they're important. For instance, fear might express that your physical integrity is under threat. So, in this case, fear discloses the value of your physical integrity, that it is important to maintain it.

However, this understanding of affect doesn't suffice to make it moral judgement. A rabbit can feel fear, but we would be hesitant to say that it is making a moral judgement in doing so. Prinz's explanation for this is that it is because the rabbit's fear is not caused by rational deliberation. Because the fear could not have been caused by rational deliberation we cannot hold the rabbit responsible for having it.

Yet narrative understanding, like Prinz's theory, does not just involve affect. How have we been thinking about narrative understanding? Narrative understanding is the capacity to think through a sequence of events in an embodied way. We have a perspective on events, in the sense that we understand how they matter to us, because this perspective emerges with our affective engagement with them.

But such thinking involves understanding that there are certain relations between parts of the story: there are patterns of cause and affect, and patterns in the roles and intentions of a character and their actions. Additionally, it is possible to articulate these relationships even if we are not currently doing so. We can understand why poor people shouldn't be given financial support if we think poor people are lazy and selfish, and financial support will encourage this behaviour. Without these connections between poverty and a character trait, financial support and the encouraging of unattractive character traits, we cannot understand the narrative. If we fail to understand the connection then we have no narrative understanding. If we question the validity of the premises but understand the connections between the parts then we understand it as a false narrative (as long as we also think it is a narrative that aims to represent the truth).

So, we should think of narrative understanding as not just an exercise of our emotional understanding but an exercise of our capacity to use a recombinant system. Storytelling and understanding involve parts, or concepts, that can come together and recombine in various ways.

We could understand these conceptual and affective capacities as coming apart in narrative understanding, or we can understand the recombinant system that we use to understand narratives as itself affective. Above I have indicated that there are reasons for thinking the latter. If we think that emotions are conceptual, and conceptual processes are those we are capable of using in reasoning, then emotions are themselves able to participate in our reasoning. On McDowell's understanding of judgements, emotions could only participate in the activity of judging if they are conceptual, that is, able to participate in our deliberative reasoning.

Charles Taylor provides a theory that makes sense of this suggestion that is, at times implicitly, driven by phenomenological considerations. He argues that the stories we can articulate co-emerge with our affective and moral sense[61]. In this discussion we will go back and forth between theory and phenomenology because they mutually inform each other on Taylor's (1985) account.

For Taylor, our strong evaluations, which emerge through our capacity to use moral concepts, are affective yet constituted through our moral language. Further, strong evaluations are the product of a narrative process, and they exist in a web of interrelated ideas and evaluations.

There is a phenomenological observation that is explained by claiming that moral concepts are affective. Which is that our moral language is experienced as already emotional, rather than leading to an emotion. Consider words like 'justice' and 'purity'. It does not seem possible to understand the concepts involved without some affective experience. To grasp their conceptual content is to feel something. If we take the phenomenology seriously, this implies that our moral concepts are constitutively affective, because to grasp the concept is, in part, to feel something.

---

[61] Taylor's theory is associated with his realist meta-ethical position. That is, he thinks that our interpretations are an attempt to articulate an ethical truth beyond what is good for a person or culture (Taylor, 2007). Prinz (2007, chapter 5), on the other hand, is a subjectivist, he thinks ethical truths are relative to individuals. However, I don't think a particular meta-ethical position is implied by the discussion here. I remain agnostic on whether our storytelling is an attempt to articulate a subject-dependent or subject-independent truth.

Taylor thinks that what distinguishes moral creatures is their capacities to make 'strong evaluations', where strong evaluations are constructed through a creature's interpretive abilities. Strong evaluations are the background evaluations through which we form particular evaluations of modes of being and living, as well as actions and events. Strong evaluations involve evaluations of what is of qualitatively higher or lower worth. And they are enabled by what Taylor (1985) calls contrastive language: language concerning what is virtuous or vicious, good or evil, pure or tainted etc. Our strong evaluations arise through the interconnections and contrasts between different moral terms, and these terms are set in a narrative understanding of what kind of life or world is of qualitatively highest value[62].

Taylor brings strong evaluations into focus by contrasting them with weak evaluations:

> *In weak evaluations, for something to be judged good it is sufficient that it be desired, whereas in strong evaluations there is also a use of 'good' or some evaluative term for which being desired is not sufficient; indeed some desires or desired consummations can be judged as bad, base, ignoble… etc* (1985, p.18.)

While positive weak evaluations may incline us to seek the object of those evaluations, we can consider these evaluations in a process of weighing up preferences. Because of this, weak evaluations allow us to reject some course of action based on contingent and circumstantial reasons, as will be illustrated shortly. Whereas with weak evaluations, there is no moral weight, in the sense of evaluating something in terms of its qualitative worth, in strong evaluations "the conflict is deeper; it is not contingent" (p. 21) because what is rejected is rejected because it is base or low.

---

[62] Note that everything said about strong evaluations is consistent with the notion that metarepresentation is not necessary for moral agency. Our emotions, while containing a pre-reflective sense of self via expressing what matters to us, do not have to involve an explicit representation of self. And when we make explicit our emotional commitments of what is important, that too can be directed at the world and need not imply that I articulate in overt terms how I, as an agent, want to live. For example, narrative understanding can involve the articulation that harmonious interactions between all life-forms are important, and can direct and justify my actions, without me making assertions about mental states.

To see this consider that we may not just have competing evaluations, for example, between flying to the Caribbean for sunshine, sea and tropical fruit, and going hiking in the highlands of Scotland for some handsome views, good exercise and a sense of achievement (similar to Taylor's example, 1985, p. 24). We may make our decisions based on some sense of which desires are morally good or bad. We choose hiking in the highlands because flying involves a qualitative judgement of lowliness of seeking gratification for selfish reasons. If our decision to go on holiday is based on a weak evaluation we might choose the cheaper holiday. In such a case, if our circumstances changed so that we had more money, then we would have gone on the more luxurious holiday. But if we pick our next holiday based on strong evaluation then we don't go abroad because we believe that flying is an unworthy action. In this case, our decisions are not changed by circumstances in the way our weak evaluations can be.

Again, I think that the appeal here is based on our experience, what Taylor is offering us is a theory that explains that experience. The phenomenology to be explained is that of our experience of a contrast between decisions where we feel no strong commitment to the alternatives other than what is practical, or what we would prefer, and those decisions where we feel called to honour some commitment of what is good or of highest worth. While we may not always honour such commitments, they come with a sense that we should do. We might see our feelings of guilt when we don't as expressive of our feeling that we have failed to live up to our moral commitments.

This background experience of some objects being of qualitatively highest worth is present even in those who try to measure everything quantitatively. They also have some background strong evaluations. Such people may "admire the mode of life in which one calculates consciously and clairvoyantly as something higher than the life of self-indulgent illusions" (1985, p. 23). If we have admiration for science above other modes or enquiry, or a commitment to (certain types of) utilitarianism, then we have an admiration of what is measurable and objective. We might think that these are the only methods that are honest and courageous, and there is something childish and delusional about other modes of life (Taylor, 2007). But then such attitudes also rely on strong evaluations: if we endeavour for

a life centred on what is rational, measurable and objective, we see this mode of life as of qualitative higher worth than an emotional, subjective mode of life.

Strong evaluations are enabled by our capacity to be narrative creatures (which, as we shall see shortly, is tied to our capacity to be interpretive creatures i.e. moral agents). The type of contrastive language that we use to tell stories is constitutive of a moral standpoint, a kind of paradigm concerning the way we view what we ought to do. It is a paradigm in that we cannot simultaneously understand both moral standpoints. When we go from understanding one narrative network to another, we undergo a paradigm shift. If I see acting compassionately as ultimately valuable, then I shun flying abroad as a selfish act. However, if I view pursuing my own satisfaction as the most worthy motive, then I might see going to the Scottish highlands for the sake of others as somewhat pompous. To the extent that it is possible for one creature to see things from both points of view, it will be by shifting between perspectives. As Taylor puts it,

> *With strong evaluations, however, there can be and often is a plurality of ways of envisaging a predicament, and the choice may not just be between what is clearly higher and lower, but between two incommensurable ways of looking at the choice.* (1985, p. 26).

What we have here is narrative understanding of the world that amounts to a holistic perspective through which we evaluate courses of action. As I will illustrate later, this perspective is narrative in the sense that it involves an understanding of cause and consequences, of how actions and events unravel and relate to each other. Our perspective situates our current predicament in a meaningful relationship to events through time.

Further, the holism of this perspective shares the structure of a story, where each part of story makes sense in terms of another part, and so different stories about what is good present incommensurable perspectives. Change enough parts of a story, and much of the rest has to shift for it to hang together as a whole. Again, this has a phenomenological underpinning: we experience a gestalt shift in our modes of evaluating, as we change our narrative understanding of what the good life, or the best world, is.

This brings us to the relation between strong evaluations and being interpretive creatures. What is experienced in the example above is that our articulation of what is of higher and lower worth emerges together with a orientation to a situation. If our articulations change, our orientation to the situation changes, and the experiential sense of the situation changes for us. To Taylor, this is the sense in which we are interpretive creatures: our articulations shape our experience: "human life is never without interpreted feeling; the interpretation is constitutive of that feeling" (1985, p. 63).

And this is another phenomenological point: we experience the world differently as our conception of it changes, and sometimes our conception changes as we try to make more distinct what our conception of the world *is*. For example, a woman might become aware of some of the effort, humiliation and unease involved in her daily interactions. Finding feminist theory a good articulation of her predicament, her daily interactions immediately strike her differently: eavesdropping on a conversation between a man and a woman in a café, she no longer hears a woman being rightly corrected by a man, but a man as patronising and dismissing a woman. In regard to strong evaluations, the claim is that our experience of what is virtuous and vicious etc. is constituted through our contrastive, and moral, language. Our language makes possible, and brings into focus, our moral sense[63].

However, "interpretation" has a richer meaning for Taylor. These articulations through which we understand the world are not formed willy-nilly,

[63] We may question this by pointing to cases where our experience of the world appears not to change even though we've articulated some novel judgement. For instance, it may be that someone comes to believe that women are not treated as human as men or as competent as them, but still experiences the above situation as one where a woman is being rightly corrected by a man. Yet I don't think it would be true that this articulation has not changed the experience of the situation. Once we recognise the possibility that the perception of someone as a woman affects our interpretation of them in certain ways, our reading of situations like the one above will be framed by the possibility that what appears to be going on is not in fact what is going on. There is a qualitative shift in how the scene is experience just by the fact that we come to recognise that it is interpreted, and so possibly *mis*interpreted. In less metarepresentational terms, while I may continue to experience sausages as delicious despite my newly articulated concerns with the meat industry, it is not that my experience of sausages hasn't changed. Now, my experience of them is a more complex affective response, of both dislike and like. Or, perhaps, I feel some disgust or anger at my love for sausages.

they are based on some emotional sense of who we are and what matters to us. Neither are these articulations mere translations of facts. They are interpretative because they are an attempt to articulate some emotionally felt truth, but that articulation can never be a perfect translation, it always leaves some questions unanswered, and room to go back and try to re-articulate. Each articulation leaves some further thing to be explained. For example, once we articulate our sense of embarrassment in a given situation, there is a question of what values and attitudes make that feeling possible[64].

Why? Because it is the nature of being language-users, that is, creatures who have a conceptual structure that allows us to ask for justifications and clarifications.

> *To say that language is constitutive of emotion is to say that experiencing an emotion essentially involves seeing that certain descriptions apply; or a given emotion involves some (degree of) insight.* (Taylor, 1985, p. 71)

Because, as language-users our experiences exist in a space or reasons, they imply the possibility that we can ask for explication or justification of them. Our moral perspective does not cause an emotion, which in turn causes action, our moral perspective exists as a web of related affective concepts, which justify and motivate our actions. So, in contrast to Prinz's position, Taylor's theory enables us to understand how an agent could act rather than just behave. For if action is characterised as behaviour that is justified by normative capacities, then Taylor's account explains how this is possible.

Narrative production and understanding is important for being interpretive creatures because it is not singular issues or acts that we try to articulate, but how

[64] As we have seen, the focus of our interpretive stories does not have to be on our psychologies. Feelings of aversion to some global event may prompt us to articulate why that event is disheartening regardless of whether we turn inwards. Watching in horror at the escalating violence in Syria, the emergence of a proxy war and dissent into a bloodbath, can prompt one to discuss exactly what is horrific and base about these events without representation of our psychologies. We could, for example, understand the badness of events by the failure of various parties to uphold their roles and responsibilities, or by the equality of all human life. Of course, many would be tempted by more metarepresentational concepts involving malicious intent, opportunism, desire for power etc. The point is that this is not necessary.

things are related. Our attempt to articulate our sense of what matters is also an attempt to tell better narratives (e.g. 1985, p. 36 – 38). In this way we try to give an explicit narrative characterisation that makes the most sense of our experience, and in making our experience clearer, we change our experience itself.

To illustrate, consider Audre, who, in the context of trying to make sense of her life tells a story about herself as a care-giver, why it matters to her, and how it has been part of her life. Maybe spurred by a moment of exhaustion, she reworks her narrative about the importance of caring about others as a narrative about how she has not taken enough care of herself. Trying to express what matters to her, in her case furthering the wellbeing of those she spends her life with, she now places herself as character towards which this sentiment should be directed. Reflecting on her past her narrative is not a simple one of acting in a caring way, but selectively acting that way towards all those other than herself[65].

It is important to recognise the dance between explicit reason and emotion that happens in these cases. Of course it may be that Audre has heard people argue that one must look after oneself, and she followed the reasoning. However, for it to be meaningful to herself depends on her felt understanding of her life. Maybe she feels unease at her sense of fragility and emotional exhaustion, which leads her to consider how she treats herself. These emotions express what her life is like at the moment, that some part of how she lives is self-destructive. The explicit inferences, once she comes to this point in her life, make sense to her emotionally as well as merely logically. They allow her to make sense of her unease and distress, her affectionate feelings towards others, and her desire to be a caring person.

This new narrative that Audre co-constructs through emotion and explicit reasoning has real implications for the direction that Audre takes her life in next, and therefore what she experiences. She learns to take her own needs and desires into account when deciding how to spend her time. Her feelings of affection towards herself grow, and maybe her relationships with others improve from her renewed energy and feelings of self-respect.

---

[65] This is a fictionalised account of the feminist theory of Audre Lorde.

Thus the narrative she weaves about herself depend on what she feels, and also change who she is. She can only articulate how important caring for people is because of her feelings of affection and love. And she only understands that something needs to be addressed in her life because her emotions direct her to examine her own wellbeing. But language is needed to organise these feelings, and to make clear to herself what her predicament is and what matters to her. In making explicit what she values, she directs her life in a particular way.

While this example involves metarepresentational narrative, not all examples do. For example, if Jake watches a movie about a rock star who always puts his career before his friends and family, then the emotional arc engendered in Jake, culminating in a sense of heartache and loneliness, may lead him to think 'close social relationships are important' and thereby affect what it is in life he attends to and pursues[66].

This analysis comes apart from Prinz's analysis of how we come to our moral sense of what is valuable and good because of the role that concepts play. Strong evaluations are constituted through a narrative network, which is a space of affective reasons. That is, narrative understanding forms a web of affective concepts that is constitutive of a moral perspective. Audre's sense of the importance of self-compassion is constituted through her story about her life. For Jake, his sense of what is important is constituted through understanding the life of a rock star.

Both Audre's and Jake's activities can be seen to be narrative in the sense that they involve understanding a context through a set of causal interactions through time. Jake, for instance, understands his interactions with people in terms of intimacy and reciprocity through time, and his affective orientation when considering whether those relationships are maintained or not.

What is of vital importance in my theory of moral judgements is that it allows us to see how moral judgements are judgements of an agent, they are judgements we can hold someone responsible for because they are answerable to,

[66] There is an interesting question here of the contribution of art to one's self and moral understanding, for relevant discussion of art, emotion and understanding, see Elgin (2008).

and not merely caused by, reason. What is meant by 'responsibility' in this context is just that some act or judgement is answerable to reason.

There is a contrast here between the framework I am working within and Prinz's. For Prinz, we are responsible for our moral judgements when they are caused by reason. However, as we saw before, moral judgements, for Prinz, do not take part in our deliberative processes. In this way, moral judgements are not really conceptual, even by his own standards.

For me, however, our moral understanding is constituted through understanding narratives that incorporate a sense of what is of higher and lower worth. When we act from a perspective constituted through understanding these types of narratives, our actions can be justified by that understanding.

We make moral judgements when we articulate our narrative understanding, and such moral judgements shape our understanding. Here, we see that in narrative understanding, the affective sense through which we understand a sequence of events is conceptual in the sense that we can draw upon this affective sense to spell out why we did something. It is in this sense that our narrative understanding justifies our behaviour and is therefore a reason for our actions.

This does not mean that it easy for us to change our deepest sense of what matters. This cannot necessarily happen without a great deal of reflection, struggle and openness (Taylor, 1985, p. 41). It is not clear that to be held responsible for something it must be easy that we can change. That is, it is not apparent that for something to fall into the space of reasons it must be simple and effortless for us to articulate our position further and examine the reasons, or lack of reasons, for it.

There is still a worry that this is a hyper-rational stance: it doesn't seem obvious that to we should only call something a moral judgement when we have reflectively reasoned about it. I will come back to this worry in section 3.4.

For now, we should note that we now have two competing phenomenological stories: on Prinz's view, our moral sense is arrived at through a rational process that is not affective. Further the deliberative processes we use do not permeate our affective and moral experience, but merely lead up to it. This is a story of two halves: one deliberative, and the other affective. If this were to

translate into a phenomenological story then our moral deliberations will be experienced as non-affective while the reactions they cause in us will be uniquely affective.

However, on a narrative understanding theory, it is through this faculty of telling and understanding stories, that we get our moral perspective which is also a moral sense. Here, we experience our moral sense as open to articulation, and we express and create this sense through telling stories about ourselves and the world. When we start questioning old stories, and start telling new stories, we experience our moral sense as being reconfigured and permeated by the stories we tell. Further, our moral understanding of the world, and the stories involved in our deliberations, are both experienced affectively.

The two phenomenological descriptions support two different theories: on Prinz's deliberation causes moral emotions, and on mine and Taylor's, moral deliberation and moral affective experience co-constitute each other.

On McDowell's account, which I am endorsing, it looks like concepts must constitute moral judgements. It is only if moral judgements are conceptual that they count as activities of normative agents, which are agents that we can hold responsible for their actions, such as human agents. And I hope that the phenomenology and theory that supports this is convincing: we *do* experience our moral sense as being open to articulation and it is changed by our articulation. It is because of this that we can be said to be responsible for our moral sense.

The shape of the argument here is not that, by developing a narrative account of moral agency, we expose a detailed critique of Prinz. Instead, the narrative account of moral agency is congruent with wider philosophical considerations, and the detail counts towards developing a robust alternative to Prinz's theory. The detail counts towards the argument first, by giving us a firmer grip on what an alternative account could look like, and second, by showing how it is an account that is coherent with phenomenological and empirical considerations.

Before I look at the empirical considerations, I want to flesh out the role of deliberation in helping to constitute a unified perspective that emerges with moral agency. I will clarify how this theory relates to issues such as moral dumbfounding,

and then return to the potential issue that this theory is a hyper-rationalist account of agency.

3.2. The role of explicit inferences in moral agency

Taylor's theory of strong evaluations and self-interpretation gives us a certain theory of how emotions and recombinant systems are co-constituents in our ability to be moral agents. Further, this is a broadly narrative account of moral agency in the sense that interpretation depends both on us being able to understand narratives, and incorporates emotion as a component that expresses our perspective or stance on what matters. Now I want to illustrate how recombinant systems are also necessary for narrative coherence by enabling coherent inferences between aspects of the narrative, and by allowing us to have an explicit sense of causal trajectories. In allowing us to consider possibilities beyond what is present now, recombinant systems are critical for us having reasons to act. Yet, language is important for moral agency in a strong sense. It is with language that we can step back and re-articulate or reconsider our narratives, so we become responsible for them in a new way. I do this to highlight and distinguish some of the roles a recombinant system and language have in our narrative and interpretive practices.

In Audre's narrative her use of a recombinant system is involved in more than just organising motivations hierarchically such that she can come to the strongly-evaluative conclusion that caring for herself is worthwhile. In addition, I think it is also being used to organise her inferential commitments coherently, and think through a sequence of events.

Maybe most fundamentally we can understand a recombinant system to be necessary for narrative because it allows one to think through the sequence of events that constitute a narrative. Human self-understanding would not be possible if we could not think through an event, its cause and its (often far reaching) consequences. Audre has far less understanding of herself if she cannot think through her past caring actions, that she did them out of love, that the consequences were detrimental to her own wellbeing.

Before I suggested that emotional cadence is necessary for MTT to be inhabited, so that MTT without the right emotional structure is not narrative. What I want to say here is that emotion and a recombinant system are both needed for narrative. While plot may need emotional cadence, conceiving of a sequence of events requires a recombinant system.

Since human narrative understanding contains some sense of what is important to us through the emotions involved in it, thinking things through narratively can potentially expose contradictions, whether we recognised them as such or not. Watching the film of the rock star, we may both envy their life because status is important to us, and feel sad at that meaningful relationships are lacking. That recombinant systems enable coherent inferences means that through understanding a story about what is a life worth living we can resolve these contradictions. When recombinant systems involve moral values we can develop a logically coherent self, with coherent *moral* commitments, we can become unified creatures, whose values and motives hang together.

Further, by 'coherence' here, I mean coherent enough at any one point for some actions to be supported by one's viewpoint, while others are not, so that we are not paralysed with indecision, as we discussed in chapter 2. This does not imply coherence through a lifetime, although, as mentioned before, continual shifts in perspective such that one rarely followed a long-term goal may be a problem for the type of diachronic agency we often associate with people.

Still, if we return to Velleman's original notion of agency, that it is the capacity to act for reasons, we can see the agency that emerges with the ability to form long-term plans as one important aspect of most human agency, but not identical to it. If agency requires us to act for reasons, and a relatively coherent, narratively-structured, viewpoint enables this, then a relatively coherent, narratively-structured, viewpoint enables agency.

Some kind of diachronicity will likely emerge with this: that is, insofar as our viewpoint remains the same, we will express this viewpoint in different situations through time. And if it is our desire for consistency that is the cause in a change in viewpoint, as Korsgaard (1989) points out, our agency is exercised regardless of changes through time. However, our narrative understanding of

ourselves in terms of roles and responsibilities will enable us to form long-term plans, that we can act on through time, and will lead to the kind of diachronic agency that Bratman and G&K are interested in. Finally, once we have metarepresentational narrative understanding, and so the ability to explicitly compare our current desires with our long-term commitments and strong evaluations, our capacity for diachronic agency will be enhanced.

However, while Audre uses language to evaluate and change her narrative understanding, I want to suggest that something more basic, a recombinant system, co-emerges with a more basic type of agency (Taylor, 1985, p. 39). This is the type of agency where we attribute someone's actions to who they are, their unique perspective. If, through understanding stories, someone comes to a unified point of view, there is a sense in which they are responsible for their actions. This is responsibility in the weak sense of being attributable to a creature with a particular perspective, a perspective we can judge as good or bad regardless of how open to change this perspective may be[67].

Because a recombinable system does not require a creature to be able to think about inferences, we may be able to attribute this type of agency to all sorts of creatures. Where creatures could be understood as having some type of recombinable system, regardless of whether this system is partial, broad-grained and context dependent (Hurley, 2003), we can think of them as acting. Whether, for instance, a lion can ever act, would depend on whether, and to what extent, we can understand a lion as being able to dissociate means and ends, and put action sequences together in different ways dependent on context.

Because such a recombinable system is affective, the perspective constituted by this system is always a value-laden perspective, in the sense that the situation is understood through how it helps or hinders a creature (see Thompson, 2007). The outcome of my proposal therefore puts constraints on what types of things we label 'agents'. While an agent need not be a biological thing, a

[67] Although, see Hurley (2003) for a discussion about this being insufficient. She suggests that you also need to have some teleological context, provided through evolutionary and/or cultural practices, that is used to determine when a rule is misapplied.

computer that cannot understand events affectively is not understood as an agent on this account.

However, without a creature having the capacity to articulate their position, and therefore question or clarify it, we could not hold them responsible in a stronger sense. This is the sense that they are responsible for the moral perspective because they could change it through reflecting on whether it was justified. We hold a creature responsible in this sense, because we understand that they could stand back and reflect on their perspective. So the claim here is that one can be an agent in a weak or strong sense. Either a creature can put together a relatively coherent position, without being able to change that position (a weak sense of agency), or a creature can be able to change one's perspective through reflection (a strong sense of agency).

That is, for agency in a stronger sense, we need to go beyond a recombinant system, and have a system that enables us to reflect on what we understand as good. I am calling this system 'language' but by this I mean any system with the function being described. Language is a system that not only allows for combination and recombination, but also allows us to stand back from combining and recombining and ask whether our attitudes are justified or unjustified. The capacity to stand back and reflect enables us to be actively engaged in forming our own perspective, to be creatures of our own making. It is when we can do this that we can expect a creature to go beyond making links between beliefs that can be justified or unjustified, to being a creature for whom it would make sense to ask them to justify their perspective or articulate it further. It is here where we start talking of a creature having conceptual abilities.

This is why language and responsibility co-emerge: responsibility is attributed to those creatures that can exist in a space of reasons, that is those creatures for whom it is *possible* for them to stand-back from their perspective to scrutinise it. Actions, now, are attributable to not just a creature with a perspective, but a creature who is responsible for their perspective. Further, *moral* agency requires recombinant systems with the type of contrastive language that enables us to make explicit what is of qualitatively higher and lower worth.

For example, Audre is engaged in the task of trying to make several different beliefs and experiences consistent with some overarching understanding of herself and her life. Prior to reflection, it made sense for Audre to direct feelings of affection and love towards others but not herself. After reflection this was shown as inconsistent with her most strongly held values. Audre has the ability to reflect on what kind of narrative is most coherent, and most in keeping with her prereflective sense of what is of highest value. Audre can be asked to defend her position, and she has a kind of control over what her position is that she lacks if she cannot stand back and reflect on it.

It need not be the case that we are constantly reflecting for us to be a moral agent, just that we can and we sometimes will, particularly if other people ask us to explain our actions. This means that we can, to greater or lesser extent, exercise our capacity to be self-constituting creatures, depending on our propensity to reflect. If someone very rarely engages in reflection then there is a sense that they have exercised their capacity to shape their own viewpoint less than someone who has engaged in reflection a lot. However, both are as responsible for their viewpoint in the sense that both could, if they wanted to, engage in reflection and change their perspective.

In this way, the type of narrative understanding we have and the type of agency we have co-emerge. Using a recombinable system co-emerges with a basic type of agency, where an act is attributable to a point of view, and having language co-emerges with a stronger type of agency, where an act is attributable to a creature who we hold responsible for their own point of view. Our capacity for strong evaluations, which involves moral languages, co-emerges with moral agency.

3.3. Clarifying narrative moral agency

There are a number of potential issues with the account of moral agency I have suggested above. For example: does it mean that one is more moral if one engages more in telling and understanding narratives? And, how does this account fit with the evidence from dumbfounding? In this section, I take the opportunity to clarify my theory of narrative moral agency.

While it is the case that narrative understanding co-emerges with moral agency, it is not simply the case that this view entails that people who tell more stories are more moral. This depends on our meta-ethical position. If one understands moral truths to not just depend on an individual, either because one is a moral realist or because one is a relativist who defines morality as relative to a culture or group, then one cannot become more moral merely by telling more stories. It also depends on whether the stories we tell are actually making available to us moral truths. That is, it is not the amount of stories we understand that matters, but the quality of our story or stories. What matters here is whether the attitudes that constitute our stories are true of the world, or your culture's view of the world.

Similarly, reflecting more will not be the only way to gain more accurate stories, although it is one way, because there may be aspects of our pre-reflective attitudes that make us more prone to notice moral issues. For example, perhaps being brought up in a caring environment enables some people to register moral facts better through the way their attention has been directed by others. In this case, some people's pre-reflective narrative understanding may be more attuned to moral facts.

Of course, if you are an individualist relativist, then the question of how one becomes more moral is a strange question. More moral compared to what? It appears that you are either moral or not. Being more or less moral may not make sense if there is no standard to compare one's own morals too. In this case, if you are capable of reflective narrative understanding then you are a moral agent and you cannot be more moral[68]. Or, a relativist could judge the extent of another's morality just based on whether another's perspective matches their own. If this is the case then, again, the extent to which someone is moral is a qualitative question about the types of stories being told, rather than the number of stories.

[68] It has been suggested that the involvement of deliberation in agency makes my account peculiar in this regard: it implies that one is more moral when one reflects more. But this cannot be the case, because Prinz's account of moral agency also relies on deliberation. His moral judgements are judgements because they are a result of agency, by which he means our capacity to deliberate.

An outcome of this account is that we now have a different explanation to the moral dumbfounding cases than Prinz presented as a reason for his sentimentalism. Moral dumbfounding is when people are presented with evidence against their reasons for holding a moral judgement, but continue to hold that judgement. For example, after reading a vignette about sibling incest, participants continue to claim it is wrong despite their reasons – e.g. "it will cause psychological harm" – not being a good reflection of the story they have been told. Since emotions and moral judgements correlate in these cases, Prinz has suggested that this is evidence for the idea that emotions, and not deliberation, constitute moral judgements.

I would like to suggest that taboos, such as incest, are cases of cultural rigid narrative understanding, that do not easily change when presented with counter-stereotypical narratives. For example, we understand incest within a cultural narrative where people suffer negative psychological consequences if they engage in it. This is important, because it is not wise to generalise from taboos to all moral judgements, considering taboos stand out as particular types of moral judgements. That is, judgements that are hard to change. So dumbfounding, when it happens, perhaps reflects a difficulty people have with understanding cultural taboos within a narrative that departs widely from the one they have been brought up with.

Nonetheless, there is some evidence that people are not always dumbfounded when presented with taboo narratives. Feinberg et al. (2012) found that if people were either naturally inclined to reassess their emotions, or were prompted to by the experimental set-up, they were less likely to display dumbfounding effects. That is, they were less likely to give reasons that contradicted the narrative they were presented with, and their emotional reactions and their explicit reasons were consistent.

So we can and do change our moral judgements when we deliberate and understand taboo behaviour within a new narrative context. For instance, if we understand a narrative where the people engaging in incest are not psychological harmed as a consequence. While it may be hard to understand particular novel narratives, this does not make it impossible.

We may explain dumbfounding the way Prinz does: that it shows that deliberative rationality does not constitute our moral judgements. However, we can adopt a more sceptical attitude concerning whether one can easily change a lifetime of learning a particular narrative through one anecdote. If, in addition to this, we also note that not all moral judgements concern taboos, and we combine these considerations with a broader empirical picture of what is going on, then we can explain dumbfounding through our narrative tendencies. As usual, the evidence underdetermines the theory, and dumbfounding can be interpreted multiple ways.

Finally, what becomes more apparent on this picture of agency is that the idea that we need a specific drive for *self*-consistency, as Velleman is committed to, is misguided. We have a drive for consistency, which includes, but is not limited to, our selves. Self-consistency is a result of us being normative creatures in general, creatures that want their beliefs to be justified by other beliefs, and their actions to be consistent with their beliefs, regardless of whether they think of their beliefs as beliefs. Or, to put it in McDowellian language, we are creatures that want to engage in the activity of giving and asking for reasons. We can see this in how narrative understanding need not involve self-reflection to be involved in having a consistent viewpoint.

3.4. Allaying hyper-rationalist fears

What remains a potential problem with this account is that it appears to imply that we only act for moral reasons, in the strong sense, when we have reflected on our reasons before that action. That is, action by an agent is defined as the result of exercising our conceptual capacities immediately prior to the action that the concepts justify. If a moral judgement depends on being able to further articulate or justify one's position then we seem to rarely make moral judgements, and then rarely act, rather than just behave (or act in some weaker sense).

However, this depends on a particular way of understanding a process as being conceptual. What I want to suggest now is that the account of (strong) moral agency above comes with a different understanding of what it means for some

process to be conceptual. On this understanding, concepts can be passively shaping our experience without them being actively deployed. What makes it a conceptual process is that one could, if they tried, clarify or justify the concepts in the moral narratives through which our moral perspective is constructed.

Before I go on to expand on how this is possible it is worth reminding ourselves of what would motivate us to take this route. The argument is that we have a host of considerations that, when taken together, suggest that the moral perspective through which we act is both experiential (an embodied moral sense) and conceptual. However this theory appears to clash with the intuition that we can act morally, and held responsible for those acts, without having reflected prior to acting. For example, if Zia – who has been known to glorify violence and declare men 'wimps' for not being up for a fight – punches Farhad because Farhad said Daniel Dennett was a poor philosopher, we would want to hold Zia responsible regardless of whether he[69] had explicitly reasoned about his actions before hand. If we want to uphold (as I do) both the theory and the intuition, then we need to explain how they are consistent.

To make sense of both theory and intuition, we can return to a formulation of perceptual experience that McDowell (2006) introduces. Here, in explaining how a perceptual experience can justify a belief, he suggests that we can understand our immediate perceptual experience as "already, an actualisation of the conceptual capacities that would be exercised by someone who explicitly adopted a belief with that content" (p. 133). So someone who unreflectively sees that the ball is red has actualised the same conceptual capacities as someone who comes to explicitly believe that the ball is red. The difference is that the latter person has *exercised* the conceptual capacities, whereas, with the former example, these capacities are being actualised without being exercised. Similarly, Zia seeing Farhad as the right target for punching is the actualisation of certain concepts pertaining to the value of violence in dealing with conflict, rather than the exercise

[69] 'Zia', while it ends in an 'a', is generally a male name.

of these concepts[70]. To see something as red, or to see someone as punch-worthy, does not require us to *do* anything: sometimes balls present themselves to us as red, and sometimes people are perceived as punch-worthy. While these are passive processes, what gives our perception of the entities that populate the world the character they have are our conceptual capacities. The claim is that our conceptual capacities are shaping how we perceive the world, without us actively engaging in conceptual practices.

This makes intelligible how someone could act for reasons without them always reflecting before they act, because "what matters is the *capacity* to step back and assess whether putative reasons warrant action or belief" (2009, p. 130). It is not required that one always does. McDowell asks us to consider a woman who sees a signpost pointing right and so turns right. He argues that,

> *What shows that she goes to the right in rational response to the way the signpost points might be just that she can afterwards answer the question why she went to the right – a request for her reason for doing – by saying "There was a signpost to the right". She need not have adverted to that reason and decided on that basis to go to the right.* (p. 130.)

We might think that this is post hoc rationalisation: maybe the woman didn't go right because she saw the signpost pointing right, she just said that is why she went right after the fact to explain her behaviour. Or, more accurately, she went right because she saw the sign pointing right, but the psychological processes that was responsible for seeing the sign and responding to it is nonconceptual.

One question here is how feasible that is: does it make sense to say that the way we respond to symbolic expressions, such as an arrow pointing in a particular direction, is not shaped by what we can talk about and reflect on? Responding to culturally mediated objects such as symbols seems to be the domain of language-using creatures. Further, in terms of our experience of arrows, it seems that before we reflect on what direction they are pointing us to, we experience them as pointing us in a certain direction. We read the

[70] It important here that 'concepts' are being used to refer to what would be needed for the stronger form of agency – that is, agency where we hold people responsible for their own perspective because they are capable of reflecting.

arrow's meaning, instead of experiencing ourselves as moving in reaction to them in the way our knee jerks if we hit the right spot. So how we immediately understand the world does seem to be shaped by what we can articulate.

Neither of these points prove that we should understand things the way I am presenting them. What I am proposing is that there is a way of explaining the role of reflection in moral agency without being hyper-rationalist or adopting Prinz's framework. Moreover, something like this account is congruent with the idea that narrative abilities can both be articulated and shaped by what we articulate.

In the context of this project, our moral perspective is conceptual in the sense that our moral sense in a particular context is an actualisation of the concepts that constitute our moral perspective. It is enough that we can reflect on it and articulate it further through further storytelling at some point. Therefore we do not have to be actively making moral judgements for our moral perspective to be a reason for action.

What is crucially different between my theory and Prinz's is that, on my theory, the moral perspective on which we act is conceptual in the sense that we can, when we want to, use it in deliberation. So, even if we do not reflect before an action, we can be asked to justify an action in retrospect. It is the conceptual nature of our perspective that allows this. On his theory, in contrast, an emotion causes our action. But, according to this view, an emotion is not something that participates in rationality, it is just an output of it. Our actions, on this account, are not merited in the sense we want them to be if they are an act of the agent, because the cause of the action cannot act as a reason in the sense of being actualisations of a rational process.

Note that the presence of reflective capacities does not necessarily imply metarepresentation. I may reflect on my belief that the world is round, rather than flat, but, as far as I am concerned, I am reflecting about the world. I could think: 'is the world round? It looks flat!'. This reflection is metarepresentational only when I reflect on my belief *as a belief*. For instance, if I think, 'should I believe the world is round? I only believe that the world is round because someone has told me so'.

We should not, then, see a theory of narrative moral agency as hyper-rationalist, because we can understand one's experience as falling in the space of reasons. This doesn't mean that we must think about our behaviour before we behave for it to count as an action, but that it must be possible that we could stand back from our behaviour and reflect on it.

4. Empirical support for the integration of concept and affect

4.1. Narratives & affect

We have explored above different facets of the argument that reasons for actions are constituted through concepts. However, the strength of the arguments above is contestable. It is controversial to suggest that talk of agency should be talk that refers to a space of reasons. And it isn't rare that people are unconvinced by theories that are supported by observation about the structure of our experience. However, the claim I want to make now is that there is empirical evidence that is congruent with my theory, but appears to be in tension with Prinz's. This means there are several lines of convergence, between conceptual argument, phenomenological description and empirical considerations, that are better explained by a theory of narrative moral agency than Prinz's sentimentalism.

There are two types of evidence I want to look at. The first concerns the congruency between the theory of narrative moral agency and a study concerning the relationship between narratives and moral experience. The second piece evidence concerns different studies on how abstract concepts are affective, which both supports the theory of narrative moral agency, and poses problems for Prinz's theory.

The first type of evidence is a study where participants read stories while having their brain scanned. This study is useful for a couple of reasons. It fits in nicely with the theory in this chapter that our capacity to work out what is of higher worth and our capacity for narrative are entangled. It also fits in with the analysis in previous chapters that the medial prefrontal cortex is important in these

processes. In general, it gives us empirical support for the idea that it is a discourse level, narrative, process that enables us to be moral creatures that act on what we value.

In a recent study Kaplan et al. (2017) showed that reading stories engaged a particular network in the brain, including the medial prefrontal cortex (mPFC). Interestingly, this network is the "default mode network", previously thought to be the resting network. Moreover they showed the same network was activated more strongly when participants read stories that included "protected values". Protected values, here, are values that are understood by the person who holds them to be values that they will not compromise on. In their study, Kaplan et al. collected personal narratives from blogs, and asked participants from the U.S., China and Iran to read them while in an fMRI scanner. Afterwards, participants were given a questionnaire. Included in it, participants were asked whether the protagonist's actions were valuable to themselves (the participant) or the protagonist. To determine whether participants understood the stories as involving protected values, they were asked whether the protagonist could be offered any amount of money to act differently than those values prescribe, and whether, if they were in the same scenario, they could be offered any money to act differently. Whether a story was seen as engaging protected values involved using the responses to both these questions.

Protected values seem to be 'strong evaluations' in that we will not compromise them. In this sense, despite the rough-and-ready nature of the probe, we can see this experiment as looking at the relationship between strong evaluations and narrative. For our purposes this study shows three crucial things. First we have good empirical support for the notion that the mPFC, which contains the ventromedial prefrontal cortex (vmPFC), is involved in narrative understanding. Second, Kaplan et al. found that the activity in the network they studied also increases as the story continues. This supports the reading that mPFC is involved in narrative understanding, considering that as the amount of information needed to integrated increases, so does the activity in the mPFC and the rest of the network. Third, we have good empirical support for the idea that strong evaluations and our capacity for narrative are closely connected. We have

reason to think that the same brain network enables both, as activity in this network is greater when we are likely to be involved in processing strong evaluations.

Other interesting points come out in Kaplan et al.'s (2017) discussion which support claims I have made in previous chapters. They recognise that the default network seems to enable mental time travel, among many other capacities, and that "interestingly, all these operations [social cognition, internally directed processing, mental time travel and self-related processing] are either involved in the processing of narratives or rely on a narrative organisation of information" (p. 5). They also review other studies that appear to show that the mPFC is central in finding and forming coherence, including coherence in stories. Further, they draw on experimental literature to suggest that the network that enables these tasks "is ideal…as a high-level coordinator across sensory, motor and memory domains" and because it is central to integrating information, it draws on different affective areas to enable "the complex emotions evoked by stories" (p. 6).

This analysis not only echoes mine in many respects, but is consistent with my description of narrative understanding as embodied and requiring one to hold together a perspective on an array of features and events. As I have suggested earlier, one way to understand this is that such a large-scale integration of so many conceptual, and experiential resources is a large part of what it takes to enable the emergence of a perspective on and in the story. While this fits well with what I am proposing, this evidence is not cited as knockdown evidence against alternative accounts. Still, if one wants to claim that this evidence is consistent with Prinz's, or some other account that differs from mine, we are owed an explanation of how it is consistent.

4.2. Concepts and affect

The empirical analysis above gives us a positive reason for accepting the idea that narrative understanding is important for moral judgements. However, there is more to be explored in the relationships between affect and concept that speaks against Prinz's account.

As we have seen, Prinz is ambivalent about whether moral emotions are conceptual or not: on the one hand, he says moral emotions constitute concepts, and on the other hand, he doesn't give them the role in deliberative reasoning he should, given his own definition of 'concept'.

I want to present some empirical evidence that gives us some reason for thinking that we should think be less ambivalent than Prinz is, and commit to moral emotions constituting concepts. Again, this is not presented to prove my proposal and disprove others. But it is evidence that needs to be explained (or explained away) when discussing alternatives to my proposal.

The empirical data below looks at the relation between abstract concepts and affect, and shows that abstract concepts are particularly affective. I think we should see moral concepts as abstract concepts. Think of what terms are classed as a strongly evaluative. Terms like: just/unjust, noble/base, courageous/uncourageous. These are abstract terms in the sense that they gather a lot of seemingly diverse events and behaviours, and classify them, together.

Further, such studies, to investigate the difference between abstract and concrete concepts, use abstract and concrete words. This suggests that the object of study are concepts in the full sense, i.e. those that we use in deliberative reasoning. Words, in the form of inner speech, are often what we use to deliberate.

Neuroimaging studies show that areas associated both with sensorimotor and emotional processing are activated when abstract concepts, of which moral concepts are a subclass, are processed (Pexman et al., 2007; Wilson-Mendenhall et al., 2011). Vigliocco et al. (2014) found that emotion-associated areas of the brain were activated more strongly in abstract words compared to concrete words. Additionally, they found that the stronger the affect, in the sense of distance from neutrality, the more certain parts of the visual system were activated.

Kousta et al. (2011) found that abstract words tended to have a higher emotional valence than concrete words. Finally, in a review of the literature on the neuroscience of abstract and concrete words, Binkofski & Borghi (2014) found that abstract and concrete word processing both share sensorimotor networks, but abstract words appear to require more affective processing and concrete words greater sensorimotor processing.

That is, the balance of affect and sensorimotor involvement, when comparing abstract and concrete words, was one of emphasis. While concrete words are based in particular sensorimotor patterns to a greater extent, abstract words are no less experiential. Instead abstract words are less based in particular sensorimotor patterns or, at least, encompass a greater degree of the general sense that apparently disparate sensorimotor patterns express. Indeed, one way to understand affect is that it requires an abstraction. Core relational themes abstract from particulars. For example, we can feel fear (that we are in danger) in situations that look different and where very different actions are suitable. If I'm scared of my boss, maybe I should report their bad behaviour to my union representative. If I am scared of that car hurtling towards me unexpectedly, I should run. In the next chapter, I will argue that the particulars of a situation are included in affect, but this will not be to dispute that affect also includes features of abstraction. Our fear is modulated by the particulars, and yet we understand both instances of fear as tokens of the same type, because there is some aspect that remains constant.

This observation about the abstract feature of affect is relevant to my project because Prinz does not give moral emotions full conceptual status. However, those concepts that we are most interested in here, concepts of what is of value or what is good and praiseworthy and which we have reason to think are involved in deliberative reasoning, are highly abstract concepts. As above, I do not want to imply that we only experience a situation as just/unjust when we reflect on it and apply a concept to it. Instead, the claim is that our experience is shaped by our conceptual understanding. So we immediately understand a situation as being just or unjust, before any reflection. Or, as McDowell would say, our experience is the actualisation of concepts related to justice or injustice. This is a visceral reaction because our conceptual understanding is affective: we experience the injustice as bodily orientation, a readiness for action.

Finally, while these studies show that abstract words are more affective and concrete words more sensorimotor, the theory in the next chapter, where sensory and affective processes are interdependent but not identical, would predict that this finding is one of degree rather than absolutes. That is, an affective understanding of our situation incorporates some particulars of what is be sensed

and vice versa. We can make sense of this by noting that, in narrative understanding, our abstract concepts are always contextualised, and so can be modulated by the sense of their context – they do not exist in isolation but in relation to other concepts and the situations that the narrative expresses. My embodied experience of injustice, while retaining some commonality, may also involve some degree of variation as the context that I am finding unjust varies. An experience of injustice that one has not been consulted on some matter is not identical to the experience of injustice at the affects of war.

So it is not just the case that, in experience, moral concepts are affective. The neurological underpinning of conceptual understanding involves brain regions associated with affect. Further, we can see why we find this more with abstract concepts, because abstract concepts rely more heavily on higher-order relations, but our affective understanding already includes such a type of understanding. Affective understanding includes an understanding of a general predicament and whether it should be encouraged or changed. Abstraction and affectivity therefore share a common structural feature.

A worry we might have is that, while such abstract affective concepts are used in deliberation, they are not used in making moral judgements in the sense that matters for Prinz. That is, they are not what we act on. To bolster this intuition we might claim that concepts cannot motivate, only nonconceptual emotions (i.e. emotions not involved in deliberation) can. Yet it isn't obvious why the conceptual nature of a processes would cause the action-oriented character of emotion to disappear. Further, some more work would have to be done to fully motivate this argument; particularly because some interpretations of the neurological literature undermine a dichotomy between cognition/conceptual capacities and emotions (e.g. see Pessoa, 2008; and Lindquist et al., 2012).

From the empirical results it appears that Prinz is misguided in his sentimentalism. We have reasons to think that moral concepts, which are concepts involved in deliberation, are affective. If deliberation is affective, this challenges Prinz's implicit dichotomy between deliberation and emotion. Further, the possibility that how we feel can be involved in our explicit cognition, seems to be supported by empirical work showing the abstract concepts are affective.

5. Conclusion

The theme throughout this chapter has been the interconnection between the capacity to reflect and emotion. Using McDowell's schema, where a psychological process is conceptual whenever we can be asked to articulate and justify it, I formulated a position that the emotional nature of narrative understanding is conceptual because we can use it to articulate our views and justify our actions. We see this can be a moral activity through Taylor's articulation of strong evaluations, where our sense of what is of qualitatively higher worth is affective, yet both open to articulation and constituted through it. This possibility is congruent with neurolinguistic evidence, where abstract concepts, such as moral concepts, rely more heavily on brain regions associated with emotions. While it may be possible that Prinz could show this evidence to be congruent with his theory too, this would at least take some argumentation from him.

In this chapter, I have explained how strong evaluations are part and parcel of our capacity as interpretive, that is, narrative, creatures. And here we have a double use of this observation, to make it clearer what the assumption is that Prinz makes that we can call into question, and the other expanding on G&K and Velleman's discussion of agency. In response to Prinz, because we can resist identifying moral sense with an isolated emotion, and instead view our moral sense as a discursive, interpretive, process. This moral sense emerges, on the view I have put forward here, through a network of affective-conceptual, narratively-related, reasons. As an expansion of G&K's theory, my view explains why emotion is part of moral agency. Not only does emotion emerge with an experiential quality that is action-orientated, it is also a constituent of narrative understanding and strong evaluations, and so enables us to make sense of, and act on, what is of highest worth.

What gives us reason to accept this account is not one argument or type of evidence, but how conceptual, phenomenological and empirical results can be accommodated by the theory of narrative moral agency, but not by Prinz's sentimentalism.

The Interdependence of Sensory and Emotional Experience

The role of the body in memory can only be understood if memory is…an effort to reopen time beginning from the implication of the present, and if the body, being our permanent means of "adopting attitudes" and hence of creating pseudo-presents, is the means of our communicating with both time and space.
Maurice Merleau-Ponty, 2012, The Phenomenology of Perception, p. 187.

1. Introduction

We have so far been considering two answers to the question: how is it possible to form an attitude about what is good and valuable? Prinz argues that, while there are all sorts of important causal components, the only constituent of our sense of goodness and badness is emotion. This is a complex claim in some ways, because while in his 2006 paper Prinz talks of emotions being sufficient for moral judgements, in his 2007 book he argues that moral judgments are a compound state that including an emotion and a representation of their object. I will explain this more fully shortly.

In response to Prinz, Gerrans & Kennett (G&K) counter that our capacity for diachronicity is central for moral judgements, specifically our capacity to re-experience our past, and imagine our future. So far, we have seen some kind of synthesis of this, in the sense that emotion has been put forward as an enabling feature of our ability for mental time travel (MTT), and narrative understanding more generally.

However, I have not yet talked about a vital characteristic of MTT. And that is that MTT appears to involve a sensory component. When we remember or imagine some place or event, there is some qualitative experience of what we would sense, if we were actually there. So G&K make sensory experience central to their conception of agency. Prinz, meanwhile, makes sensory experience important to moral judgements insofar as it is part of the representation of the object that forms a compound state with an emotion.

In contrast to both theories, it looks like I give no role to sensory experience. When explaining my conception of narrative understanding, I have

mainly talked about how emotion is involved. Unlike MTT, it isn't obvious why sensory experience should be part of narrative understanding. However, in this chapter I argue that sensory experience must be part of narrative understanding, because the integration of sensory and emotional experience is fundamental to the perspectival sense that constitutes it.

In the end, I argue that G&K's more embodied conception of moral agency must be right. But that does not require that I adopt their stance that MTT is what is crucial. Instead I submit that narrative understanding always does, to some extent, include sensory understanding. But to understand why sensory experience is involved in narrative understanding, we must reject Prinz's theory of the relationship between emotions and sensory experience. Because it is only through the interdependence of sensory and emotional experience that we can understand how they co-emerge with a perspective. "Perspective" here is taken as one way to understand how narrative understanding is embodied and action orientated, as will be explained at the end of this chapter.

While this argument counters Prinz's claim of the independence of emotions and sensory experience, it also feeds into my positive proposal about how to best understand moral agency in the following way:

i) Moral agency entails being able to act for reasons;
ii) Narrative understanding is necessary for acting for reasons;
iii) Emotional experience is a constituent of narrative understanding;
iv) Emotional experience is interdependent with sensory experience;
v) Therefore, sensory experience is integral to moral agency.

To make this point about the role of sensory experience in agency, however, I will first argue that a sensory-affective amalgam is constitutive of a subject feeling present in the present. The claim is that our feelings of being present in narratives, which is what I will claim is also our sense of perspective in them, are constituted in a similar way. Much of the chapter, therefore, is concerned with our capacity to have a perspective now.

So this discussion about the interdependence of sensory and emotional experience has a dual-purpose with respect to the thesis dialectic. First, it questions the relationship sensory experience has to emotions and moral

judgements that Prinz proposes. For him emotions constitute the moral aspect and sensory experiences are conjoined to that. Second, by establishing sensing and feeling as co-constituting each other, it establishes and explains why sensory experience, in the service of narrative understanding, is central to moral agency.

To establish the interdependence of sensory and emotional experience, and how this enables us to feel present in a world, I start with Merleau-Ponty. He provides a transcendental argument establishing this interdependence, which can equally be understood as 'inference to our current best explanation'. I use this as a starting place to explain how, if we examine our everyday experience, this interdependence is evident. Neuroscience corroborates the theory that sensory and emotional experience are intertwined. Similarly, despite clinical psychology apparently providing counterexamples, not only is it consistent with the neuroscience and the phenomenology, but the phenomenological account provides us with novel insights into what is going wrong in cases of psychosis. In all, I think my account, compared to Prinz's, more accurately describes our phenomenology, allows more consistency between the disciplines, and allows greater insight into certain psychological conditions.

Then I return to narrative understanding. I argue that the sense of our imagined situation co-emerges with our sensory-affective access to it. Through the embodiment that constitutes narrative understanding, we pre-reflectively experience an identity between our perspective in narrative understanding and our everyday perspective. So the affordances we experience in narrative understanding are available in the present. Our current perspective, and therefore our ability to act, is inseparable from our status as narrative creatures.

2. Independence v.s. interdependence

2.1. Prinz's independence argument

Lets begin by looking at what Prinz's claim that emotional and sensory experience are independent consists in, and his motivation for this claim. Prinz argues that while emotions are representations of the relationship between organism and

world (e.g. 'fear', for Prinz, represents that one could be harmed), sensory experience is not a constituent of emotion, but is a partner to that feeling. He states that,

> *While there is a sense in which emotions are directed at particular events, that does not mean that they represent those events… The events are represented by mental states that combine with emotions. When I am sad about the death of a child, I have a representation of the child's death and I have sadness attached to that representation* (2004, p. 62.)

In the case of this theory, emotions, and the representation of the particulars of a situation, are essentially distinct. They are conjoined, rather than inseparable. If this is the case, it leads us to expect that there should be some cases where either sensory experience is highly distorted and emotion is experienced normally, or emotion is experienced unusually but sensory experience is experienced normally. The independence of two processes implies that neither are reliant on each other, at least for their core functions, so that one process can be impaired while the other one isn't and vice versa.

Prinz frames this issue in terms of the 'formal objects' of emotion versus the 'particular objects'. The formal object is designated by the core relational theme. Fear has an intentional object, it points towards danger, but this object is its formal object. The particular object, for example, a lion, is an add-on to the emotion, not a part of it.

Prinz makes this move for several reasons. First, it is not clear what preserves different instances of the same type of emotion if emotions contain their particular object. For example, if emotions include their particular object, then it is not clear what makes fear the same type of feeling in each case. Sometimes fear is caused by a tyrannical boss, sometimes it is caused by an exam, and sometimes it is caused by lions. Prinz argues that what makes these uniformly fear in all cases is that they are concerned with their formal object, that these are all threats to our safety. As Prinz puts it,

> *One can generally find a common theme behind the range of things that elicit any given emotion. Consider a number of things that might cause sadness: a child's death, a report on the political crises in the Middle East, a divorce… They are alike in one respect: they all involve the loss of*

> *something valued… Emotions represent their formal object, not their particular object.* (p. 61-2.)

Furthermore, Prinz objects to the general framing that occurs when the particular object is seen as included in the emotion. To see emotions this way is, for Prinz, to see emotions as secondary qualities, that is, as an experience enabled by the property of an object. But Prinz protests that, unlike colour, which is a secondary quality, we experience an emotion as within, rather than part of the world: "the feeling of rage, for example, is not projected onto the object of rage; it is experienced as a state within us" (p. 61). Because emotions differ from colour in this way, Prinz argues that emotions cannot be secondary qualities. This quote also indicates that Prinz understands his account to make sense of our phenomenology: because emotions are interoceptive states, they are experienced as within, and because sensory experiences are exteroceptive states, they are experienced as without. The independence of the psychological processes of emotions and sensory perception is mirrored in how our experiences present themselves to us.

Finally, Prinz argues that emotions are not secondary qualities because emotions are not strictly speaking response-dependent. They have a formal object that represents a state of affair independently of including a particular object:

> *I can continue to think about the death after my sadness has subsided and I can continue to be sad after my thoughts of the death subside. The mental representation of an emotion's particular object can be doubly dissociated from the emotion it elicits* (p. 62.)

So, Prinz thinks the emotions and sensory experience are separable. But this idea has an interesting relationship to his idea that emotions are sufficient for moral judgements. This is the claim he makes in his 2006 paper, but it is not the claim made in his 2007 book. Remember, from chapter 1, that Prinz thinks that moral judgements are a compound state composed of an emotion and a representation of their object. Here he claims that, when we form the moral judgement that 'pickpocketing is wrong',

> *The anger that arises in the observer is not simply free-floating rage. The anger was triggered by the experience of pickpocketing, and it gets bound to the representation of the pickpocketing. The result is a compound state: anger at pickpocketing. This compound constitutes the judgement that pickpocketing is wrong…* (2007, p. 96.)

Here, the particular object is included in the moral judgement. It is a constituent and not cause in that the moral judgement is comprised of anger and a representation of the pickpocketing. A representation of the pickpocketing has to be included in the moral judgement, because for Prinz, the moral emotion only has a formal object. In the case of moral anger, the formal object is either injustice (when it is indignation), or the violation of rights (when it is righteous anger) (2007, p. 70). Because the moral judgement is, in this case, that "pickpocketing is wrong", and Prinz argues that emotions do not include the particular object, moral judgements must be a compound state that includes not only emotions, but also a representation of their particular object.

Therefore, Prinz doesn't have a straightforward sufficiency thesis as long as the representation of the pickpocketer includes sensory experience. Nonetheless, it is the emotion, in the moral judgement, that provides the moral appraisal of the situation. The sensory component, if there is one, fixes the particular object of that moral sense. So it is part of the judgement in so far as the particular object is part of the judgement, but it is not the moral part of the judgement. What I present in this chapter challenges Prinz in two ways. It challenges his theory of emotion, as just involving the formal object, and it challenges the role he gives sensory experience in moral judgement, because it proposes that sensory experience co-constitutes our moral sense, and is not just conjoined to it.

I do not seek to undermine Prinz's assertion that emotions have formal objects, and to some extent this explains what is similar between different instances of a type. Nonetheless, it's not clear that our fear of different things (our boss, an exam, a lion) isn't, too some extent, altered by what that thing is. That is, it is not clear whether emotions being understood, in part, by their formal object excludes them also being understood, in part, by the particular object. In response to his second and third objection, I think he has simply got the phenomenology wrong. Emotions do seem, to some extent, to inflect our experience of the world, and furthermore, the world, to some extent, seems to impact us bodily. That is, the distinction between colour being 'out there' and emotion being inside is less absolute than Prinz suggests. Similarly, our feelings of sadness at death will

incorporate some sensory imagery, and certain imagery to do with the death of a loved one generally will incorporate some sadness.

Instead of diving into a detailed response to each of Prinz's arguments now, what I want to present in section 3.1. is a robust phenomenological case for thinking that sensory and emotional experience is interdependent. This will give me the tools to express more fully what is unsatisfying with Prinz's proposal. Then I go on to look at what the neuroscientific and clinical psychological evidences suggests. As explained in my second chapter, the phenomenology and science mutually support and make sense of each other, by providing a constitutive and enabling story that fit together. Further, Prinz's theory is not supported by the evidence, and makes it opaque how the phenomenological and empirical evidence fit together. First, however, I want to say a bit more about the shape of my claim.

2.2. Defining the alternative

In contrast to Prinz, I will be arguing that emotional and sensory experience are interdependent. Note that I am not stating that emotional and sensory experience are identical. Instead, they overlap, and both are enabled because of this significant overlap. This leaves open the possibility that there are some neural processes that are more associated with emotional experience than sensory experience and vice versa[71]. In this chapter, I argue that my theory is preferable because it is more accurate, enables different evidence to be understood as consistent, and it explains more.

However, my claim may seem odd, given my habit of still using the descriptions 'emotions' or 'embodied appraisals' and 'sensing'. How can I use these descriptions while concurrently claiming that they mutually constitute each other? The answer is that we can see different descriptions, at least on the surface, as picking out different characteristics. Yet these descriptions are not picking out entirely different phenomena. There are different ways to characterise significantly overlapping phenomena. Very roughly, we can think of sensory experience as 'experiencing an object'; emotional experience as 'experiencing some

[71] See Pessoa (2008) for a similar observation about the relationship between cognition and emotion.

meaning'; and experiencing action affordances as 'the capacity to understand how, and to what extent, I can engage with, and/or manipulate, an object'. In claiming that all these activities co-constitute each other, I am claiming that none of these things come apart completely. So, on this picture, it is through the mutual constitution of each that *perception* emerges. Unlike some uses of the term 'perception', I do not understand it as not reducible to, or synonymous with, non-affective sensory experience.

The phenomenology shows that, when we pay attention to our experience of the world, emotions and sensory experience are part of each other. Prinz, however, claims these experiences are distinct. So my theory my accurately describes our experience. The neuroscience is consistent with the phenomenology, and shows that the processes that enable each characteristic are constituted partly by processes that realise the other characteristics. And the symptoms of mental health disorders can be fruitfully explained by this theory, so clinical psychology is enriched through the account that I present in this chapter.

To be able to talk about the relationship between sensory and emotional experience I will be making use of the words 'interoception' and 'exteroception'. Interoception is the perception of the state of one's body, including sensing the state of one's organs, changes to one's breathing rhythm, muscles, joints, and feelings of comfort and discomfort (Barrett & Bar, 2009). The theory of emotion I have relied on, that of emotions as embodied appraisals, depends on interoception being constitutive of emotion, since emotion is the feeling of the body that expresses what is of value, or what is meaningful, for us. However, 'interoception', as I will be using it, is as a slightly broader category than bodily experience, since by 'experience' I mean something that we are conscious of. Interoception will refer to both nonconscious information pertaining to, and conscious experiences of, the body.

Exteroception is similar to what is often understood as the five senses, that is, those senses that tell us something about the world beyond the skin. Similarly I use exteroception to include non-conscious processes.

I want to show that both emotional and sensory experience rely on the combination of interoception and exteroception. The phrases are important to pursuing this project because one of the claims is that when we experience a situation, nonconscious processing of the body and world have already occurred and been integrated.

The use of 'interoception' and 'exteroception' also helps to clarify the way that emotional and sensory experience may be interdependent rather than identical. These phrases allow me to leave open the possibility that sensing doesn't directly involve all of the interoceptive processing that enables emotion, and emotion doesn't directly involve all of the exteroceptive processing that enables sensing[72].

The theory of interdependence is important for my overall project in two ways. Because it challenges the role Prinz gives sensory experience in moral judgements. And because it is precisely because of the interdependence of sensory and emotional experience that narrative understanding has a perspectival phenomenology and, with it, the capacity to motivate action. So, in this chapter, I develop the theory of perspective that began in chapter 3.

To begin to argue for these points, I start with the phenomenology. And once we have this in place, we can begin to review Prinz's arguments.

3. Converging evidence

3.1. The phenomenological perspective

3.1.1. A transcendental argument

The phenomenological perspective offers two, interrelated, types of arguments to articulate the nature of our experience. The first is Merleau-Ponty's transcendental argument. The second are observations about everyday experience that take the

[72] This is in line with Pessoa's (2008) account of the neural connectivity, where some regions of the brain are 'hub' regions that cannot be reduced to traditional categories (in the case of his paper he is disputing the cognitive/affective distinction) while more peripheral areas are dedicated to more specialist function. See, Lindquist et al. (2012) for a similar conclusion about the neurological underpinnings of emotion.

form of first-person data. These two arguments are linked for me, because the end of the transcendental argument provides the starting point of my articulation of the relationship between sensory and emotional experience by providing a new framework for understanding experience, one that undercuts the assumptions Prinz needs for his characterisation of phenomenology. If the transcendental argument works, then my characterisation of the relationship between sensory and emotional experience must be broadly correct. Nonetheless, for some readers the transcendental argument may arouse suspicion. For them, the more straightforward characterisation of experience may be independently convincing. So I present these two arguments separately. I start with introducing Merleau-Ponty, and his transcendental methodology.

One of Merleau-Ponty's aims in his *Phenomenology of Perception* (2012) is to fully capture and explicate the structure of human perceptual experience. In doing so, he engages in transcendental phenomenology (Gardner, 2015). He wants to explain how it is possible that we come to experience ourselves as a subject within a world of objects. These conditions of possibility are, for Merleau-Ponty, simultaneously phenomenological and metaphysical. Before we experience ourselves as subject among objects, we have a prereflective experience that is pre-objective and are able to explain how it is possible that we come to understand the world objectively. Such prereflective experience comes with a metaphysical claim that we are not fundamentally subjects or objects, but a type of thing that transcends that distinction. As Gardner argues, for Merleau-Ponty this is the starting point from which we should articulate our experience. Our account of our fundamental and prereflective experience should neither place us as an object amongst objects, nor as pure subjects.

Abstractly, the method that Merleau-Ponty uses to lead us to this position involves setting up a dialectic between a position he calls 'empiricism', and another that he knows as 'intellectualism'. The former is the thesis, and through the failure of the thesis, we come to its antithesis. This fails too. And so we are led to a synthesis that in some way combines notions from both thesis and antithesis, but also rejects a common assumption. However, in taking something from both thesis and antithesis, a dialectic also seeks to transform the notions used in these

frameworks. While empiricism seeks to tell us about experience through the causal powers of the object, and intellectualism from the constituting powers of the subject, Merleau-Ponty's synthesis, if done right, should transform our very understanding of what it takes to be an object or a subject. This dialectic seeks to lead us through various ways of understanding experience in a logical journey, so that the dissolution of each step leads to the next, and the resulting synthesis articulates and clarifies a truth that is self-evident, but that we can only fully grasp at the end of our dialectical journey. Having summarised Merleau-Ponty's general method, I now turn to his particular dialectic concerning our capacity to have sensory experience.

One commitment that Merleau-Ponty takes to characterise empiricism is the claim that we can understand sensory experience through the way that objects affect the organism. Conversely, intellectualism makes the claim that the organism creates sensory experience. Empiricists have a purely passive view of sensory experience. For them, experience is explained just as a chain of causal events from the object to the organism, where the organism is an object in space that obeys the laws of nature. Intellectualists, on the other hand, have a fully active view of sensory experience: it is the creation of a thinking thing that posits external objects.

As usual, Merleau-Ponty's view is a middle way: sensory experience can only be understood if we can dissolve the theoretic divide between object and subject. Because both give partial stories neither empiricism nor intellectualism can explain sensory experience. While intellectualism and empiricism differ in whether they take sensory experience to be an active or passive process, they also share a fundamental assumption that Merleau-Ponty seeks to undermine. Both take an objective grasp of the world as fundamental. In doing so they get things the wrong way round: we cannot appeal to the objective world to explain sensory experience, because it is the perspectival nature of sensory experience that explains our ability to grasp the world objectively. In terms of empiricism, we can only come to know facts about the world through how the world effects us, so empiricism ignores what makes their explanation of sensory experience possible. Merleau-Ponty uses the example of understanding a cube in objective space. He

notes that our primary way of understanding a cube is not by the "cluster of objective correlations" (2012, p. 210) between the system of moving body and cube. Merleau-Ponty writes:

> *Now, if the words "enclose" and "between" have a sense for us, they must borrow it from our experience as embodied subjects. In space itself, and without the presence of a psycho-physical subject, there is no direction, no inside, no outside. A space is "enclosed" between the space of a cube as we are enclosed between the walls of our room.* (Ibid.)

Our primary experience is of ourselves as bodies that are related to objects external to us and it is through our experience that we can understand the concepts of empirical science: because we have a perspective on the world we have the concept 'between'. 'Between' is then something we come to understand through our embodied experience, and then apply to objects in objective space[73].

The intellectualist flips the empiricist's thesis, and tries to understand experience through the activity of a constituting ego. However, intellectualism shares an assumption with empiricism that Merleau-Ponty rejects: "it remains true to say that intellectualism also takes the world as ready made" (p. 215). The subject of perception is an all-powerful creator of a world of objects, which are set out before us perfectly and in its entirety.

Through reflecting on experience, intellectualists come to claim that there is an ego that creates the cube and sustains the relationship between the cube, the body, and the sensing mind:

> *The entire system of experience – world, one's own body, and the empirical self – is subordinated to a universal thinker, charged with sustaining the relations among these three terms. But since this universal thinker is not engaged in the system, the terms remain what they were in empiricism, namely, causal relations laid out on the level of cosmic events.* (p. 215.)

Because the cube is known to us through the constituting power of thought, we know it completely, and its relation to us is fully grasped.

Crucially, this story contradicts our phenomenology. If intellectualism were right, and we constitute the world, then our experience of the world and ourselves would be complete and perfect. Instead,

---

[73] For a good example of what our bodies contribute to our experience, see Merleau-Ponty's discussion of depth in section B, chapter 1, part 2 of *Phenomenology of Perception*.

> *We have the experience of a world, not in the sense of a system of relations that fully determines each event, but in the sense of an open totality whose synthesis can never be completed.* (p. 228.)

According to Merleau-Ponty when we attempt to explicate the experience of sensing, we see that intellectualism must be wrong. The world is not ready-made or even ever fully made, it is partially made and remade.

For Merleau-Ponty (2012), sensory experience does not come apart from our ability to act in the world and vice versa. He claims we have a sense of our body due to what actions are possible: our body is experienced in terms of what is virtual, what it could do, rather than as "a thing in objective space" (p. 260). To have a sense of our body in space, however, takes more than just action tendencies. It also involves what Merleau-Ponty calls a "spectacle", the situation through which my actions unfold. Spectacle, here, refers to the environment one experiences oneself as immersed in[74]. Merleau-Ponty suggests that,

> *My body is geared into the world when my perception provides me with the most varied and clearly articulated spectacle possible, and when my motor intentions, as they unfold, receive the responses they anticipated from the world.* (p. 261.)

We inhabit the world when we experience ourselves as being immersed in a spectacle that we have a perspective in and on. And our perception of a spectacle before us, when it becomes clear(er), is itself experienced as an enticement to act in certain ways. Perception involves being poised towards a spectacle, it is a certain bodily attitude. But the experience of action affordances is integrated with "the invitation of these very gestures" through our capacity to sense (*ibid.*). Both are necessary for the feeling of our body as part of a world[75].

---

[74] I take it that the use of "spectacle" indicates a scepticism towards talking of the objects of perception in any purely mind-independent way.

[75] Merleau-Ponty's phenomenology therefore involves a particular metaphysical position. For him, we constitute the world, but partially. Because of this, the external world, as mind-independent, is absent from his metaphysics.

We may wonder whether Merleau-Ponty achieves the non-problematic metaphysics he seeks. For it was meant to be a problem of both the empiricist and intellectualist that they took the external world for granted, and yet talk of the 'world' slips into Merleau-Ponty's framework again and again. Maybe we can bypass this problem if Merleau-Ponty

What we see through Merleau-Ponty is that experience cannot start though objective awareness because the experience of objects presupposes our experience of embodiment. Importantly, when we consider intellectualism we see that neither can our experience be of a constituting subject, because such subjectivity entails the same objective world that we've come to realise is presupposed by a more fundamental experience. It just is not the case that, in our pre-reflective experience, we have a distinct or all encompassing grasp of ourselves or objects. This leads Merleau-Ponty to the conclusion that our pre-reflective experience precedes, and makes available, our experience of a subject as discrete from the world. Our pre-reflective experience is body-involving insofar as it is spectacle-involving, our bodily engagement includes a sense of the thing they are orientated towards. We are led to this conclusion by the failures of empiricism and intellectualism, and we end at a point that seems to make better sense of our phenomenology.

I take this to be a transcendental argument that establishes that body and world are entangled in pre-reflective experience. Or we could say that experience depends on an integrated intero-exteroceptive process[76]. This implies that affect, normally thought of as interoceptive, and sensory experience, normally thought of as exteroceptive, are interdependent.

It should be noted that while I have described the form of the argument as transcendental, in the sense Merleau-Ponty has provided proof that only a particular conceptualisation allows us to understand how experience is possible, we can also understand this argument (or transcendental arguments in general) as 'inference to our current best explanation'. In this case, Merleau-Ponty has

---

doesn't mean 'world' as a reference to the external world, but to the totality of experiences. For our purposes, we can see his argument as partly tackling how we ought to understand the character of our experience and put broader metaphysical issues to one side.

[76] I take reference to intero- and exteroceptive process to be a more accurate, though less poetic, way of understanding how our experience is constituted, because it does not rely on a simple dichotomy between body and world. A simple dichotomy is not one that Merleau-Ponty wants to ascribe to, since our bodily orientation is world-involving and our experience of the world is body-involving. The use of 'interoception' and 'exteroception' enable us to explain how that is the case more clearly: experience of our bodies involves an integrated intero-exteroceptive process, as does our experience of the world.

provided the best explanation so far of how experience is possible, and this explanation is open to revision.

Prinz, however, may find this description unproblematic. After all, emotions and sensory experience must be causally connected for him if emotions are to represent what a situation means for an organism. However, it should be remembered that emotions, for Prinz, are independent from sensory experience insofar as they do not contain a particular object. As we shall see, this is the sense that I am implying that emotions and sensory experience are interdependent. A question remains, however, concerning how much Prinz can agree with my analysis while maintaining that we should understand sensory experience and emotional experience as separate but closely causally related rather than interdependent. I shall come back to this worry later.

In what follows I try another, although related, track to Merleau-Ponty's argument. And that is a simple characterisation of experience that I hope rings true, and therefore undercuts Prinz's assumptions about why emotions and their particular objects must be independent. The characterisation is enabled and implied by Merleau-Ponty's philosophy, but cannot be straightforwardly read off from it, since his comments that I think are most related to my endeavour do not explicitly mention affect.

3.1.2. Characterising experience

Merleau-Ponty (2012) argues that any kind of perception follows this pattern of the interdependence of body and world. What is sensed is a "communion" (p. 219) between the spectacle and a bodily attitude, so that a situation that is "about to be sensed poses to my body a sort of confused problem", for which, if it is to be elucidated, my body "must find the response to a poorly formulated question" (p. 222). So, sensing follows a pattern of the (partial) resolution of ambiguity, when the confusion of our experience prompts a bodily attitude that has a better grip on our situation. This grip is constituted through the co-emergence of sensory experience and experience of action affordances. Note, that even before we resolve our situation, our experience is sensori-affective: our body relates (us) to the

ambiguous situation *as* a problem, *as* a confused question, that is, *as* something meaningful.

A particular example is that of perceiving ourselves and other objects in space:

> *The spatial level is, then, a certain possession of the world by my body, a certain* hold *my body has on the world… It sets itself up when, between my body as the power for certain gestures… and the perceived spectacle, as the invitations of these very gestures and as the theatre of these very actions, a pact is established that gives me possession of space.* (p. 261, original emphasis.)

The perception of space depends on both the experience of being enticed by a spectacle, and our body as being able to find a pattern that fits with this enticement[77].

One way to phrase this is that, to inhabit a world, interoception and exteroception must become interconnected. For experience depends on both a perturbation experienced as originating from the spectacle (i.e. exteroception) and an embodied reaction to it (i.e. interoception). The interconnection of these two poles enables the experience of being within a world that we move through. Our feeling of presence in the world is the feeling of a body orientated to a situation, and our situation as presenting a meaningful milieu for us. A milieu that presents itself as to be engaged, or not engaged, with, and that impacts us from the start. Our bodily orientation already incorporates what is external, and what is external expresses a sense, which is already relational to us. Which is all another way of saying that what we sense is affective, what is affective is sensory, and this interdependence is crucial in constituting our perspective in the world.

This intero-exteroceptive process manifests in general experience as the interdependence of sensory and affective experience. The world before us speaks

---

[77] The description of how we inhabit space is particularly important for my project. Here we have a description of what it takes to be immersed in an actual situation, and it depends on the coming together of our experience of our bodies with "the perceived spectacle". Since narrative understanding is my topic, my ultimate concern here is how it is possible for someone to experience a situation that is non-occurrent. I will argue later that there is an important continuity between our current experience of our situation and our capacity for narrative understanding.

to us. It is affectively coloured through the affordances it presents. It entices us, or challenges us, facilitates our action or gets in the way. It is easy, and we sink into it, or it rejects us and attacks us. What is sensed always has some significance, expresses some values. The objects we see, hear, smell, evokes from the start, a sense of themselves, which always includes a sense of how we are related to them. The smell of coffee in the morning, when we experience it, already presents itself as rewarding. And the grey sky, wind battered trees, and rain lashing on the window has a mood. There is no perception without a sense of the perceived as somehow related to us. That is, there is no perception that does not contain affect.

While it may be clear that weather and coffee are affectively experienced, one may worry more about events and objects that are less obviously emotionally charged. What about the experience of looking at a white wall, or the surface of the floor? It seems likely that affective experience here is less intense. Yet we can make sense of these situations if we understand sensory experience as being narratively structured, in that our sensory experiences are constructed, in part, through an emotional cadence.

In this vein, our particular experiences are meaningful through the narrative context that frames them. Looking at something relatively homogenous or familiar is a sensorimotor activity where consecutive actions are coupled with the sensory experience one would expect. The affective experience of homogenous walls and floors co-occurs with the experience of ease or normality, in that what they mean for our sensorimotor activity is predictable. This is already a kind of narrative because there is a sequence of events that are understood through our expectations and both being easily met, and our activity being repetitive. Exactly how this is experienced affectively depends on the particular narrative context. One may be expecting more stimulation, in which case our sensory experience is integrated with the experience of boredom, or one may be cherishing some reprieve in a hectic day, in which such activities the homogeneity is perceived as peaceful. If we think about minimalism design, this supports this observation. The motivation behind minimalism is not that we will stop affectively experiencing our surroundings once they are simplified, but that they will be experienced as peaceful (or open in some way).

This is the same for our experience of more obviously affective sensory experience. Dark clouds and rain may be experienced as awesome (in the traditional sense) rather than gloomy if one is wrapped up in warm waterproof clothing, and the sky sets off a majestic scene during a walk in the Scottish highlands. Our sensory experience in this case too is partially constituted by an affective cadence that situates our current predicament inside a larger narrative of the grandness and power of nature.

In common cognitive science parlance, we can put these affective experiences simply in the language of affordances. The world already, from the beginning, is something that we have a bodily orientation towards. But this bodily orientation is not pure mechanics, it couldn't be an orientation if it was[78]. Our experience of the world is somewhat integrated with our experience of the possible actions available. But the sense of the world that comes with those possible actions has affective weight. We might say that perception is always atmospheric, even when that atmosphere is cold and clinical, when we feel cut-off and distant, there is something meaningful about our perception.

Similarly, we can see affective experience as incorporating something of what it points towards. A bodily orientation makes sense as an engagement with some thing. Our poise is not (only) towards an abstract sense of what we are related to, but a poise towards particular objects. It is the thing that has a sense for me, that I am directed towards. My feeling of affection towards the smell of coffee incorporates the particular bodily orientation I have towards coffee. The feeling is inflected with the sense of its object, and is not simply a conjunct to it. My melancholy at the dour Edinburgh weather is not the feeling of an abstract state of affairs, somehow conjoined to an object, but is bound up with what is sensed[79].

This brings into view something that is troubling in Prinz's account. I am arguing that it is not right to characterise our affective experience as involving only a formal object. Instead the particular features of a situation are part of our

[78] That is, an orientation points beyond itself, but this cannot obviously be captured by a mechanical explanation, where one things leads to another.

[79] Merleau-Ponty's argument above is meant to bring us to a similar point to this, but through a different argument. The language of affordances can be understood as one way of explaining what Merleau-Ponty means by the body being geared towards the world.

embodied orientation, through the actions they afford. When I fear various objects: an exam, a boss or a lion, that fear expresses that I am in danger in each situation, but my precise affective orientation also incorporates the particulars of the situation. That is, my emotional experience of fear is not homogenous but inflected with a sense of a particular situation. Partly because what something means to us is far more nuanced than core relational themes allow: I fear the lion because what is at stake is my physical integrity; I fear my boss because they have the power to damage my social status and emotional wellbeing; and I fear the exam since my intellectual credibility, and long-term career options, are being called into question. While the formal object can explain what stays constant in feelings of fear it does not explain my precise feeling. The particular object makes an important contribution to defining what matters. These differences in the situations show up through inflecting the emotions experienced and comes together with differences in my mode of interaction in these spheres, a bodily engagement that expresses the particular way that a situation presents a danger for me.

One could respond that emotions must be defined by their formal object. It is through doing this that we can understand what is held in common between different instances of the same emotions. However, if one wanted to use the commonality between emotions to argue against emotions having particular objects, one would have to argue that having a formal object and having a particular object are mutually exclusive, which Prinz has not shown. What I am arguing here is that emotions are characterised by both their formal and their particular object.

Further, Prinz's analysis of secondary qualities as being experienced as outside, while emotions are experienced within, doesn't seem so absolute when we consider our experience. When we consider our experience, objects are emotionally meaningful for us. The dark clouds express gloom, the smell of the coffee is enticing. Barrett et al. (2007) make a similar point about what they call 'background core affect' – our ongoing implicit affective states that shape our experience of particular moments:

> *Background core affect is experience as a property of the external world rather than a person's reaction to it. We experience some people as nice and others as mean, some food as delicious and others as distasteful, some pictures as pleasing and others as negative.* (p. 388.)

So it isn't that emotional experiences are experienced as fully within, instead features of the world are also experienced as expressing some affective characteristic. And, on the flipside, our sensory experiences are partly experienced as action affordances.

These partial crossovers of our situation as experienced as affective and involving a bodily orientation, and our affective bodily orientation as being directed towards the world, is integral to the unification of our experiences. This is why Prinz's account of sensory and emotional processes as distinct and separable components of experience is unsatisfying. For experience does not present itself as the addition of distinct parts, as though our experience of a particular situation has decipherable components like the bricks of a building.

It is this brick like structure that presents a problem for Prinz in explaining how experience situates us in a world. We can see this as another failure for Prinz's theory to account for the phenomenology. A collection of experiences does not add up to an awareness of ourselves in the midst of a situation. On the contrary, if our experience was as Prinz describes it – where one part of our experience is an internal experience of an abstract situation, and a separate part is our experience of an external particular situation – then experience is of non-integrated segments, and the feeling of being a subject is absent. To feel present is to experience our bodies as pointing beyond themselves towards the situation we find ourselves in, and we experience that beyond as communicating meaningfully to us, directing our embodied stance in particular ways. For there to be a world before us at all, as Merleau-Ponty notes, is through a communion of body and world. The feeling of presence in the world therefore requires that sensory and emotional experiences co-constitute each other.

This observation about how perspective emerges serves several purposes. It is central to the thrust of the chapter, yet it may not seem obvious because of the way my arguments criss-cross each other. It is part of the framework that

undermines the possibility that emotion could be, by itself, sufficient for our moral sense, by explicating how it cannot, by itself, exist. As we shall see, this is a central point for explaining certain features of clinical psychology. Finally, it will be central to explaining the phenomenology of narrative understanding, and explaining why that phenomenology co-emerges with our capacity to act on such understanding. That is, it helps explain how we can act for reasons that are understood narratively.

Prinz could suggest that his theory need not apply to our phenomenology. His distinction between emotional and sensory processes may not be a distinction that is present in experience. The problem with this is, as will be explained below, that we also have neurological reasons for thinking that sensory and emotional experience is interdependent. However, he could still maintain there are good philosophical reasons for treating them as distinct, even while accepting all the evidence I am offering. He could, for instances, accept what I am describing as a good way to understand how the emotional part and the sensory part of a moral judgement are linked. In that case, I suggest that the tight integration of such a link warrants us to understand sensory and emotional experience as not independent, but interdependent. Nonetheless, if Prinz still prefers to think of these experiences as separate but tightly linked, then the disagreement is semantic, which I would be willing to accept.

3.2. The neuroscientific perspective

We have seen that Prinz's theory doesn't fit with our everyday experience. Yet, maybe, what shows up in experience is not sufficient for the characterisation of our psychology. It could be that a neurological description of emotions and sensory perception shows them to be independent, but our experience of them is integrated. However, this route is blocked for Prinz. The observation that emotion and sensory experience are enmeshed, fits with a review of neuroscientific literature by Barrett & Bar (2009). That is, the neuroscience is consistent with our experience.

Barrett & Bar (2009), in their review of the brain activity involved in perception, conclude that, "an affective reaction is one component of the prediction

that helps a person see the object in the first place" (p. 1331). "Prediction" here is a reference to the predictive processing theory of cognition (Clark, 2015). In this theory, our brain's activity is constantly using past patterns of activity to predict the present input. Input, if judged[80] to be salient information, is used to update the prediction, or is ignored. Perception happens when the predictions more-or-less match the input. So we can understand Barrett & Bar to be claiming that sensory experience depends not just on predicting exteroceptive information but on interoceptive predictions too. They base this conclusion on the role of the orbitofrontal cortex (OFC), which is thought to be crucial for (visual) perception. The OFC is thought to integrate information from interoception and exteroception to enable a judgement to be formed on the value of a given situation for an organism.

Barrett & Bar (2009) explain that neuroscientific evidence points towards the *medial* OFC being the basis of basic 'gist-level' information concerning the relevance of a situation to a person's wellbeing and appropriate potential action. This is integrated, through strong reciprocal connections, with visual information concerning where an object is (i.e. the visual 'where' pathway). These are quick brain responses that allow the body to be prepared for a situation before objects or events are consciously perceived.

The *lateral* OFC contains multimodal neurons that receive detailed information about visual features (from the visual 'what' pathway) and interoceptive information from the anterior insular, which is thought to be important for conscious emotional experience. These processes are slower. And Barrett & Bar think they generate rich conscious multimodal perception that includes the affective value of a situation. The affective value can either be attached to the object of perception, or a subject's reaction. You may experience the lion running towards you as scary, or experience yourself as joyful in response to the birth of a relative.

[80] We shouldn't understand 'judge' here in the everyday sense. What is being proposed is not that there is an intentional being weighing up input and deciding what is important. Instead, through mechanical biological processes, some input is able to change predictions while other input is ignored.

However, according to Barrett & Bar we should not understand the fast and slow processes as two different types of sensori-affective prediction. Rather, they are likely to be one dynamically evolving process, so that gist-level processes and awareness of what an object is are mutually forming. This is because there are reciprocal connections among all the brain areas involved: the medial OFC receives some high resolution visual information, and the lateral OFC receives some gist level visual information; the medial OFC and lateral OFC are connected through intermediary brain regions; and the visual 'where' and 'what' pathway are strongly interconnected. Hence, evidence from the activity in OFC suggests that perception is constituted via dynamic sensori-affective processes.

Other neurological evidence also points to the idea that it is right to see affect as constituted through exteroception. Barlassina & Newen (2014) review evidence that suggests that disgust is associated with brain regions that integrate interoceptive and exteroceptive information. While the interoceptive information related to disgust is associated with the posterior insular, disgust itself is associated with the anterior insular, which combines both interoceptive information with information from multimodal sensory regions.

Further, if we understand emotion not just to refer to the state of the body, but also an expression of the relationship between body and world, then Barrett & Bar's (2009) evidence also suggests that exteroception is involved in emotion. It is the medial OFC, where interoceptive and exteroceptive information meet, which is responsible for the "initial affective information about what an object might mean for a person's wellbeing" (p. 1329). This is based on various neurological evidence showing that the medial OFC guides autonomic, hormonal and behavioural responses to an object, all responses that are involved in affect (e.g. Barbas et al., 2003). Furthermore, the lateral OFC integrates "bodily information with sensory information from the world to establish an experience-dependent representation of an object in context…This conscious percept includes the affective value of an object" (p. 1330). It is not merely that sensory experience is affective: conscious affective experience is enabled by intero- and exeteroceptive information being integrated. So the neural enablers of emotion, in both the medial and lateral OFC, are not associated only with interoception, but with the point where interoceptive

and exteroceptive information meet. Hence affective experience is as dependent on exteroception as sensory experience is dependent on interoception. What comes in to view in experience is a sensori-affective situation.

This is consistent with Merleau-Ponty's claim about how we have a perspective on the world, because a perspective, for Merleau-Ponty, depends on the integration of body and world in experience. Insofar as we think that phenomenology and neuroscience act to mutually inform and constrain each other, and that their mutual agreement acts as an indicator of theory success, this supports the theory of emotions and sensory experience as interdependent. Again, this evidence may not logically contradict Prinz's theory. If he wants to maintain that sensory and emotional experience of independent, yet concede all the evidence, then the remaining dispute is a semantic debate which, as mentioned above, I am not contesting here.

One thing to note in both the phenomenology and the neuroscience is the relationship between action and emotion. Both these disciplines suggest that disambiguating what we are facing involves the experience of how one is poised to engage with the world. Such poise is affectively construed, our bodies are orientated, that is, meaningfully pointed towards, the world.

Similarly, Barrett & Bar mention that,

> *The medial OFC not only realises the affective significance of the apple, but also prepares the perceiver to act – to turn away from the apple, to pick it up and bite it, or to ignore it.* (p. 1330.)

The wording Barrett & Bar use suggests that the affordance and the affect are independent, but we can look at the same evidence another way. To see, to feel, to act, are totally interdependent. The affective significance of an apple involves action affordances. If Snow White were to have seen the green slime dripping off the apple, the fear ignited would involve the affordance of avoidance, and this bodily attitude would have been part of her seeing that she has been given a poison apple. So we can see that action affordances contain a sense of what it is we are interacting with and vice versa.

This is beautifully consistent with Prinz's observation that emotion simultaneously is a perception of a situation and being poised to act in a particular way. Evaluating what we are facing means intuiting how we could act. This will become relevant when discussing the best way to understand what is going on in psychological disorders.

The account I have developed here also overlaps with a theory of affectivity developed by Colombetti (2014). For her, affectivity is part of all sense making. Sense making, for her, is an organism's activity of inhabiting and bringing forth the world around them, which includes how it relates to the situation. Affectivity, as a creature's sense of the significance of their situation, is characteristic of all sense making. By considering both empirical and phenomenological evidence, Colombetti also concludes that the appraisals involved in affect are embodied and action-orientated, and may be tightly interlinked with sensory processes, where these are different characteristics of the same process rather than different component parts. That is, Colombetti argues that these are not sequential parts of a process, where for example, we make an appraisal, and then we feel something, or we feel something and then we prepare for action.

3.3. The psychological perspective

3.3.1. Depression & psychosis as intero-exteroceptive disturbance

We have seen that the neuroscience and the phenomenology are consistent with each other, and the phenomenology is accurately described by my theory that affective and sensory experience are interdependent. However, it may seem that clinical psychology presents an alternative picture: one where emotional and sensory experience are independent.

The theory that sensory and emotional experience are interdependent, rather than independent, predicts that there should be no clear cases of changes to affect without changes to sensory experience and vice versa. If sensory experience and emotional experience are not intertwined, we would expect there to be possible cases of extreme sensory distortion without emotional distortion.

On the face of it, clinical psychology appears to support this reading. Mental health problems where there is a huge disruption to our normal affective

experience, such as major depression, are not obviously associated with sensory disturbances. Similarly in psychosis and schizophrenia, which are generally thought of as mainly disorders of cognition and sensory experience, this may not be associated with a distortion of affective experience. However, I do not believe that these cases pose the challenge that they appear to: depression does co-occur with distortions to our sensory experience, and psychosis and schizophrenia co-occur with huge changes to our affective experience. Moreover, I think the claims I made above, concerning a sense of perspective emerging out of the integration of intero-exteroceptive processes is borne out by this data. We see this because what I am claiming is that a general disturbance in this process correlates with a distortion in our sense of being in the world. I'll look at depression and psychosis in more detail in turn.

While depression is often thought of as a distortion of affective and embodied experience, many also report a change in how the world appears, although this is quite a subtle change. People who have had experience of major depression often talk of how the world seems both the same and different (Ratcliffe, 2013). Further, people with depression report that it is hard to draw apart bodily complaints of feeling clumsy, numb and un-coordinated from changes in the world. This appears to support the argument that our interoceptive experiences, our sense of what our bodies can do, and our perception of the world are all integrated. Along with low mood, lack of motivation, and anxiety, people experience the world as "bereft of any positive enticement for action; it no longer draws one in and thus seems distant, detached, not quite there" (Ratcliffe, 2013, p. 585). As the body changes its relation to the world so does our perception of the world.

Although we could interpret the reports above as relating only the changes in emotional response rather than a more general change to perception, given the discussion of phenomenology above, I think that the best way to interpret this is as a disturbance in perception due to a disturbance in the ability to interact with the world. As Merleau-Ponty notes, the presence of the world before us requires a capacity to find a bodily attitude that makes sense of a perturbation. If one's capacity for embodied appraisals is impaired, then the theory of interdependence

predicts that one's perception of the world will also change. Merleau-Ponty's analysis of how it is that we are normally able to feel present in the world makes it possible to see reports such as the one above as trying to express something literal.

One particular expression of someone with depression highlights how their affective state was constitutive of their experience of the world. He said, "the shadows of nightfall seemed more sombre" (Styron, as cited by Ratcliffe, 2013, p. 586). Rather than describing feeling sombre while looking at shadows, it is the shadows themselves that seem sombre. This, along with Ratcliffe's larger project concerning the phenomenology of depression, indicates that how we feel and how the world appears do not come apart.

Another reason to take these claims literally, that they are about one's experience of the world and not just one's self, is the continuum between affective disorders (such as depression and anxiety) and psychosis. Armando et al. (2010) found depression and distress correlated with bizarre experiences and perceptual abnormalities. Similarly, Wigman et al. (2012) found that people with anxiety and major depression were more likely to display at least one psychotic symptom compared to those without a mood disorder. Obvious disturbances in sensory experience do seem to come in tandem with mood disorders. To frame these findings in terms of the theory I've presented here, the descriptions of subtle sensory changes are on a continuum with more obvious sensory disturbances. Full-blown hallucinations do not have to be seen as completely cut-off from more subtle changes in sensory experience.

On the flip side, psychotic episodes have generally been understood as disorders of sensory experience and cognition. Delusions and hallucinations are the typical characteristics associated with psychosis. For example, someone might have the delusion that there is a government conspiracy against them, or might hallucinate people talking to them. We should care about psychosis, because it potentially shows that one can have a distorted sensory experience without a change in emotion. In turn, this would show that sensory experience is independent of emotions.

Yet, there is increasing awareness that psychosis is a thoroughly affective experience too. Because of this, psychosis cannot be used as a counter-example to

a claim the sensory and emotional experience constitute each other[81]. Vodusek et al. (2014) found that people in the prodromal stage of psychosis, full blown psychosis and post-psychosis all had unusual experiences of emotions. They 'experienced more negative and disturbing emotions than controls' (p. 254). Similarly, people who do not currently have psychosis, but are members of groups that are at high risk of having it in the future have altered affect, including anhedonia, intense emotions, and emotional confusion (Phillips & Seidman, 2008). These groups were: people who were first-degree biological relatives of someone with psychosis, people who were in the prodromal phrase[82], and people who scored high on schizotypy[83].

Yet this might seem suspicious: now the possibility arises that relatives of people with psychosis, those in the prodromal phrase, and people with schizotypy have altered affect without altered sensory experience. It is not clear that this is so. People who are prodromal often do report perceptual[84] disturbances (Yung et al 2003; Subotnick & Nuechterlein, 1988). Similarly, a feature of schizotypy is that one has unusual perceptual experiences (Claridge et al. 1996; Mason, Claridge & Jackson, 1995). And it would not be implausible that family members of people with schizophrenia have perceptual disturbances considering that they reported unusual affect, and people who have mood disorders experience the world as strange, and are more likely to have hallucinations than those who don't.

So while psychosis is often characterised by changes to sensory experience, this doesn't seem to happen in isolation from changes to emotional experience.

---

[81] Yet, neither is it the case that evidence from co-occurrence of psychosis and emotional disturbance straightforwardly supports for the constitutive claim, for that I am relying on the neurological and phenomenological evidence.

[82] The prodromal period of psychosis refers to changes in people's experiences and behavior that falls short of being understood as full-blown psychosis (Yung & McGorry, 1996) and may or may not develop into psychosis (Yung et al., 2003).

[83] Schizotypy scores depend on the theory that everyone has aspects of their personality that lie on a continuum between non-psychotic tendencies and experience and psychotic tendencies and experience. People who score high in schizotypy have tendencies that are more like those of people with psychosis than average, compared to the rest of the population.

[84] These studies tend to talk of 'perceptual' rather than 'sensory' disturbances. I retain their original wording here, but note that many people take 'perception' and 'sensory' to be analogous.

While this fits neatly both with the idea that sensory and emotional experience are mutually forming and with a Prinzian view that emotional and sensory experience are robustly causally connected, the evidence from neuroscience and phenomenology gives weight to the former interpretation. Furthermore, unlike Prinz's theory, mine also explains why a disturbance in sensori-affective processing co-emerges with a disturbance in our sense of being in the world, as we shall see.

3.3.2. Localised versus global disturbance

However, some clarification needs to be added. The theory that sensory and emotional experiences are interdependent predicts that disturbances in one will not be found without disturbances in another. But, understood simply, this is clearly not the case. People who have very different sensory capabilities don't necessarily have emotional impairments: think of people who are blind, or deaf. Similarly, large changes of mood that fall short of counting as pathological do not appear to cause sensory disturbance. Strong anger because of your stressful commute to work does not prevent one from seeing the computer in front of you when you get to your desk.

To get a better hold on this, we need to return to the theory that Merleau-Ponty puts forward, and our neuroscience supports. Here, we are present in the world if our capacities for interoception and exteroception can work smoothly together as one intero-exteroceptive process, such that our bodily feelings can find a way to tune into the spectacle. But our bodily feelings are action-orientated, they are involved in understanding possibilities for actions. Similarly, our perception of the world depends on us understanding what it means for us, including how we can meaningfully act in it. It is through the integration of interoception and exteroception that we feel present in a meaningful world. Affect, significance, sensory perception, and action-orientation come together on this account. Note that the claim is not about specific intero- or exteroceptive senses, that is, not about sight or proprioception, but about interoception and exteroception in general, how they come together, and how they are involved in the experience of action affordances.

Impairments in particular senses, then, should not be a problem, as long as there is some interoceptive and some exteroceptive component to be integrated such that one feels immersed in a world, which is experienced as an enticement for meaningful action. Blindness and deafness are therefore not a problem; there is still some exteroceptive input through other organs. Whatever exteroception is available is enough to form an embodied appraisal of how we are related to the world, which is also a signal of what actions are most suited to the situation. For example, a blind person uses tactile and auditory processing, together with interoception, to get a sense of how they are positioned spatially and affectively in relation to other things. What is important is that they have some partial functional counterpart to vision.

Similarly, it is not so much particular changes in mood that are a problem in sensory experience, but the general affective impairment. Major depression is characterised as a mood disorder, and a meta-analysis of psychological literature by Bylsma et al. (2008) found that people with major depression had a decreased affective response to stimuli compared to controls. Although both negative and positive affect was reduced, positive affect was reduced more. Heller et al. (2009) found that neurological responses were consistent with the theory that positive affect is attenuated in major depression. The subtle changes in the way the world seems in major depression can therefore be understood as a change in sensory perception due to an impairment in relating to the world in a meaningful way. The world is distant because our capacity for meaningful action orientation towards it has decreased and so it isn't as present to us. As Merleau-Ponty suggests, space is *lived:* what surrounds us is not just geometrically arranged objects, but also an arena of objects constituted by how graspable and interactable they are. That positive affect is reduced more is telling: positive emotions often invite us to interact, so a generalised reduction in positive affect comes with a decreased sense that objects are calling to us.

Unlike extreme emotions outside of mood disorders, major depression is a wide-spread change in our capacity to form embodied appraisals at all, particularly positive ones. However, one may note that even particular strong emotions can change our perceptions of the world. For example, feeling extreme indignation and

envy watching your current crush flirting with your best friend may increase the resolution and intensity of every gesture, and shorten the apparent the distance between you and them. Similarly, when feeling anxious in an everyday way, I experience light as brighter and noises as louder and more sudden.

We can also understand psychosis as what occurs when there is a large and global disruption of intero-exteroceptive processing, resulting in a highly distorted experience of body and world. Remember that we are trying to understand the distinction between some particular deficiency or disturbance to interoception or exteroception, and a global disruption to one of these processes. A global disruption, I am claiming, will impact both sensory and emotional experience, in a way that a particular disturbance, such as a single strong emotion or loss of sight, will not. Like depression, I think psychosis is another instance of this more global disturbance.

De Haan & Fuchs (2010) analysed qualitative data from structured interviews with two people who had experienced an episode of psychosis. What is most striking is the metamorphosis to their prereflective experience of being in the world. The first theme they identify is a loss of self. This includes a loss of self-coherence and a lack of meaning and motivation: "You cannot just get up and do something. Nothing means anything to you. I simply cannot assign myself; I don't know what I want to do, what I am doing, who I am" (p. 329). This also includes a disturbance to the skin/world boundary: "my skin is extremely thin" "I am too sensitive…I think I cannot defend myself" (*ibid*.).

The participants with psychosis also described their feelings of estrangement and detachment from their bodies and the world. This came with metaphors of being a machine, and expressions of disorientation that accompanied it: "In general, I didn't have a sense of my body…my face became increasingly strange", "the world is not tangible any more…if you cannot be part of it, the world automatically feels different", "I feel as if I am sitting on some distant planet and there is somehow a camera in my head and those images are sent there" (De Haan & Fuchs, 2010, p. 329), "I look around… and I'm dizzy, all is like a machine" (*ibid*., p. 330,).

To understand this in terms of the theory above, the feeling of inhabiting the world requires the co-ordination of interoceptive and exteroceptive processes, such that we can perceive coherently and experience our bodies as our selves, and our selves as embodied actors. Psychosis, as a global disruption of the intero-exteroceptive process disturbs the normal perspectival quality of experience leading to disturbances in experiences of self and world. Co-emergent with this confused processing, one gets disturbed emotion, including intense emotion, anhedonia and confused emotion[85]. The disruption to our intero-exteroceptive processing disrupts emotion due to emotional experience being constituted through this process. The strange sense of self and world experienced through psychosis comes together with this disturbed emotional capacity.

As seen above, all these experiences of self, world, and emotion come together with a lack of meaning and motivation. Not only did the participants in these interviews feel unmotivated, they also talked of their effort in acting, and their clumsiness: "There were periods in which I felt extremely badly coordinated…I found myself to be extremely clumsy, somehow, when walking" (De Haan & Fuchs, 2010, p.330). To act, participants either had to pay very close attention to what they were doing or, as one participant explains, "switch off, my mind was totally away from my body"[86] (ibid). That is, their capacity to engage meaningfully, purposefully and fluidly in the world, to have a pre-reflective sense of themselves in action, has been undermined. This is exactly what one would expect given the current framework because psychosis occurs not when one loses a particular sense or has an angry outburst, but when there is an overarching disintegration or intero-exteroceptive processing, which co-occurs with a disordered capacity for emotion, sensory experience and action.

---

[85] Vodusek et al. noted that people with psychosis were "less able to discriminate between different emotions"(p.2). I think we should see this inability to discriminate as not an epistemic failure, but as reflecting that emotion in psychosis can be more indeterminate because of the dysregulation to the intero-exteroceptive process.

[86] Note that such automatic behaviour may indicate very minimal, or absent, prereflective awareness. That patients with psychosis operate more successfully when they switch off (almost) completely, is congruent with the theory that disruptions to intero-exteroceptive processing co-emerges with disruptions to our sense of perspective.

It should be noted that Vodešek et al. (2014) found that what is described above is more characteristic of pre- and post-psychotic experience. Full-blown psychosis came with "being at one with oneself" (p. 5), feeling intense positive emotion, being highly motivated, experiencing oneself as in control and experiencing the world as deeply meaningful. This transformed eventually into a threatening, alienating experience. Again we can see how disturbance of an intero-exteroceptive process can be seen enable such an experience: here an intense increase in positive affect is co-emergent with increased action-orientation, and a revitalised experience of self and world.

It is true that co-occurrence does not prove the interconnectedness of sensory and emotional experience. I take that to be shown by phenomenology and the neuroscience, rather than the clinical psychology. Nonetheless, what I take the clinical psychology to support is the phenomenological claim that our embodied perspective on the world arises through intero-exteroceptive processing. If one buys the interdependence thesis then, one gets an explanation for why disturbances of self co-emerge with distorted sensori-affective processing for free, because this process is, to some large degree, what constitutes our sense of being in the world.

However, the case of Ian Waterman may seem to disprove the theory that interoception is necessary for sensory experience. Ian Waterman suffers from a loss of proprioception (Cole, 1995; Meijsing, 2000), meaning he lacks the normal immediate sense of how his body is organised that most of us have. Proprioception depends on nerves that register our muscle, tendon and joint position. Yet, Ian Waterman has no problem with feeling emotion or perceiving the world. He did, however, have major problems moving initially, although he learned to do so through paying close visual attention to where his body was situated. This initial problem co-occurred with a strange sense of self.

The first thing to note here is that proprioception, while an inner sense of our body, is not all there is for interoception. These two senses are often seen as distinct, although here they cannot be entirely – both are a sense of our body, and so fall under what I am understanding as 'interoception'.

Because Ian Waterman still has other interoceptive capacities, he retained his ability to sense the state of his organs, and he could feel deep pain, fatigue, hot and cold. These remaining capacities are integral for understanding how one is related to the world, because they are important for forming embodied appraisals. If this, as I have argued, is part of sensing objects in the first place, then we can see how loss of proprioception would not wipe out the capacity to sense the world. Ian Waterman has the capacity to form embodied appraisals.

This is in contrast to psychosis and depression. In these cases, while it initially looks like there is a localised problem, what we found was that there were pervasive perceptual problems. This, I was suggesting, could be explained through the general attenuation of intero-exteroceptive processing or an impairment in the co-ordination and integration of intero- and exteroceptive processes. It is a general disturbance that explains the disturbances in the sense of presence that accompanies depression and psychosis. However, the claim is not that people with depression and psychosis have no perceptual experience, and that they have no sense of presence, just that these are both markedly altered.

With Waterman, his issue is confined to one type of experience: he does not have a pervasive perceptual disturbance, but a problem localised to one modality. Unlike Waterman, people with depression and psychosis cannot use other modalities to compensate, because, I am have argued, the problem is not confined to one of their modalities. Waterman is more similar to someone who lacks a particular modality, like someone who is deaf or blind.

Still, some puzzles remain. Another function I have given to interoception is to co-ordinate with exteroception to give us a sense of our position in the world. So, I've given interoception two (partially conceptually distinct) functions: being a key part in our capacity for embodied appraisals, and being central to the experience of being immersed in the world, both working through the interaction with exteroception. A joint and co-ordinated intero-exteroceptive process is needed to enable the experience of self and world to come together, since our selves take a stance towards some situation, and the world is experienced through our embodied appraisals. However, if someone cannot adopt this bodily stance,

then my theory may make it hard for me to explain how they could feel present in the world and act.

Crucially, embodied appraisal already contains a readiness for action, it is a bodily stance insofar as it is an attitude towards the world. Waterman feels present because having an action orientation does not require currently having the motor control needed to act.

One way to understand this is that we don't need to think that embodied appraisals only rely on our feeling of our body as it is. Neuroscientist Damasio (1999) also talks about feelings as partly based on our sense of how our body could, or ought to, be given the situation we are in. Damasio calls this the 'as if body loop'. According to this theory, emotions rely partly of our sense of our body as if it were in a particular state[87]. We may not actually be cowering when we face an object that threatens us, but might nonetheless experience cowering as appropriate for the situation. This works as another explanation of why Waterman can sense the world prior to regaining his motor abilities: being able to experience oneself as if one were in a posture is a partial substitute for being in that posture. Just as we feel fear in dreams without cowering, so can someone who cannot move easily. Because even without cowering, we can experience the situation as one that calls us to cower.

However, there is more going on here than that. Ian Waterman's visual exteroception partially compensated for some of the functional roles normally played by proprioception. He could move about only granted that he was constantly looking at his body and making a conscious effortful attempt to move it (Meijsing, 2010).

We can say that Waterman has a kind of *visual* proprioception, and that his experience of inhabiting the world still depends on a capacity to co-ordinate the

[87] Colombetti (2014) raises some issues with the existence of a neurological 'as if body loop', since much of the biological realisers of emotions appear not to be brain-based. However, as long as some neural circuitry is involved, this may be part of our current affective experience even while such an experience is possibly attenuated or altered by not being accompanied by the normal non-brain realisers. Further, some of the experience of, for example, a situation calling us to cower, may not be based on an 'as if body loop' that is enabled by the brain, but actually occurring bodily changes, such as hormonal changes, that we refrain from acting on.

way he can move with information about his surroundings (Meijsing, 2010). That is, we can say that one of the functional roles normally played by proprioception can be, to some extent, played by particular exteroceptive information. Since, to move, Waterman has to engage in constant and conscious visual attention, we can see that he cannot totally make up for the loss of proprioception. Yet, Ian Waterman does not threaten my theory if we see the phenomenology as offering a functional description of what is needed to feel present in the world, if we remember that other interoceptive processes retained by Waterman play much of that role, and if we have some flexibility concerning how people with impairments use other capacities as partial replacements (similar points are made in sensory substitution literature, see Noë & Hurley, 2003).

It is through this capacity to act that Waterman regained a normal sense of inhabiting the world through his body. Before he learned to move in this way, Waterman says he felt dead. Similarly, other people who have lost proprioception but did not learn to move have related to their body as a pilot in a ship they cannot really control (Meijsing, 2010). One thing we can take from this is that although embodied appraisals involve action tendencies, and give us some sense of self, our capacity to be the author of actions does change how we experience ourselves in the world. Large impairments in our capacity to act change our experience of self, although action affordances are enough for some (distorted) feeling of a self in the world. Loss of proprioception did not remove the understanding Waterman already had of affordances, and the relationship between movement and perception, so his perception was enabled by a working co-ordination between extero- and interoception.

We started this line of enquiry through the observation that exteroception appears to be able to fail without harming our experience of affect, and we seem to be able to knock out interoception without damaging our ability to sense the world. I have argued that, in the cases where this appears to be so, only parts or aspects of interoception or exteroception are missing. When there is a partial impairment other processes can play a sufficiently similar functional role to the capacity that has been lost. What is more of a problem for our sensing and affective capacities is

when there is an impairment of our general intero-exteroceptive processing. This results in odd experiences of self, affect, and world. Evidence from clinical psychology is therefore congruent with a theory of the interdependence of sensory and emotional experience. Furthermore, Merleau-Ponty's theory of how we feel present in the world can help explain the experience of people with psychosis, as well as when, and why, loss of proprioception will disrupt of sense our being in the world.

3.4. A good theory?

We have now seen three different perspectives concerning whether emotional and sensory experience are entangled: phenomenological, neurological, and psychological. This matters because if emotions and sensory experience are entangled then it cannot be only emotion that constitutes our moral sense, and it is not right to argue that sensory experiences are part of moral judgement only insofar as they provide the particular object of the judgement. On my theory, sensory experience is a constituent to what is moral about our perspective, rather than distinct from it. Further, this theory of the relationship between feeling and sensing will form a part of my explanation of how narrative understanding is possible.

What I want to do now is summarise how the relationship between phenomenology, neuroscience and psychology relates to the two different theories being compared. The theories being compared here is the theory that emotions and sensory experience co-constitute each other (my theory), to the theory that they are tightly causally coupled, but not constitutive of each other (Prinz's theory). I am not aiming to conclusively prove that my theory is stronger. Instead, I aim to show that we have some good preliminary reasons to favour my theory over his.

Crucially, my theory is more accurate than its rival. My theory makes sense of our daily experience, that sensory experience is affective and vice versa, while Prinz's theory doesn't. In addition, if one follows Merleau-Ponty's transcendental argument, then it looks like there is a conceptual reason to suppose that only a

theory where interoception and exteroception are integrated could explain our capacity to perceive (see section 3.1.1.).

My view also enables greater consistency between disciplines. It allows three domains of enquiry to fit together. On the other hand, the view of independence can only be understood to be accurate with regards to clinical psychology. But on this theory, the three disciplines I have looked at become inconsistent. That is, the clinical psychology is now inconsistent with the neuroscience and the phenomenology.

The consistency of different disciplines can be counted as a theoretic virtue once we share commitments similar to those that Wheeler (2013) proposes. On this theory, the insights of one discipline can help us discern potential strengths and weakness in another. When the focus of each discipline is on the same subject, we might think of them as offering different perspectives on that subject. Where perspectives converge on what characterises our subject matter, we have more reason to be confident in our characterisation of that subject matter. Where there is disagreement, each discipline has the task of re-evaluating its methodology and conclusions.

Therefore, we should see the sense of being situated in the world as requiring the entanglement of motor intention, emotion and sensory experience. Sensory experience is one way that we hook on to the world in the here and now, including by it being involved in our emotional experience. And emotion as an essential way that we are connected to the world and motivated to act, incorporates both interoceptive and exteroceptive experience and is an expression of how organism and world relate to each other.

However, while much of the clinical psychology could be made sense of through either theory, only my account enables the novel insight from phenomenology to elucidate what is going on in the examples from clinical psychology. The phenomenological support for the theory of interdependence also presents an intero-exteroceptive process as being a major constituent in our sense of inhabiting a world. Here, then, we have an explanation for why disruption to the integration and co-ordination of interoception and exteroception co-emerges with

disruptions to one sense of embodied self, like in psychosis[88]. Another way to say this is that it looks like my theory has greater explanatory scope than Prinz's. If intero-exteroceptive processing co-emerges with a sense of perspective, then, my theory, unlike Prinz's, can explain cases of distorted perspective.

So, there is good reason to believe that sensory and emotional experience co-constitute each other. They are not independent. Regardless of whether one is fully convinced by the transcendental argument, one can see this theory as fitting closer with our daily experience, and as creating a framework that shows different disciplines to be consistent. Additionally, the characterisation of experience that this theory is consistent with offers novel insight into particular psychological disruptions. That is, the package that my theory offers appears to have greater explanatory scope than its rival.

The result is that sensory experience is given a different functional role in our moral perspective than the one that Prinz gives. While on Prinz's account sensory experience may enable a moral judgement to have a particular object, it is not part of the affective moral sense itself, but conjoined to it to constitute a judgement. That is, on Prinz's theory there is a division of labour between sensory experience and emotional experience within a moral judgement. The emotion provides the moral sense, and the sensory experience provides the particular object. On my theory we see a subtle shift. Here, both are co-constituents of each other: sensory experience partly constitutes our moral sense itself, and emotions partly constitute the particular object of that sense.

4. Sensory Experience and Narrative Understanding

I started this chapter by mentioning that there are two reasons for looking at the relationship between emotion and sensory experience: to present reasons for doubting the role Prinz gives sensory experience in moral judgements and to

[88] It is worth noting that a correlation between depression and an altered sense of self has also been found (Lambert et al. 2001).

present a reason why narrative understanding, like MTT, must require sensory experience. This latter aim blocks a possible objection to my account of moral agency. This objection would go as follows: MTT, but not narrative understanding, requires sensory experience; sensory experience plays some crucial role in moral agency; therefore MTT is important for moral agency in a way that narrative understanding is not.

Moreover, the above discussion contributes to my positive proposal for moral agency by contributing to an explanation of how narrative understanding enables us to act. In the phenomenological analysis above, I have so far explained why sensory experience must be necessary for us to experience ourselves as present in the present. What I want to do now is look at the implication of this in regards to how we inhabit narratives, our past and future, and act in the present. That is, what does the theory above, concerning how it is that we experience ourselves in the world, have to do with narrative understanding and moral agency?

I have argued before that narrative understanding emerges when we have the capacity to have a perspective on a sequence of events. What does this have to do with moral agency? Well, narrative understanding constitutes our capacity to have a sense of what is most valuable to us, because it enables us to form a coherent set of interrelated affective reasons through which we understand the world. And, I have claimed, these reasons motivate.

What I want to suggest now, is that this narrative process, with its perspectival character, depends on our sense of being a self within the world presently. I gave a first approximation of how that is possible in chapter 3. There I argued that emotion, as an embodied appraisal, was important for having a perspective at all. I noted that emotions have a bipolar structure, that they place us on one end of a relation. Because emotions are embodied this placing is felt. And emotions, as embodied appraisals, tell us how that thing we are related to matters to us. Lastly, this involves an embodied stance, since having this sense of how an object matters to us is also understanding a set of possibilities for action. Now we are in a position to see that emotion alone doesn't get us this.

Importantly, we have seen that sensori-affective experience co-emerges with the experience of finding ourselves amid the world, which is concurrently experienced through the possibilities of action that are afforded. Exteroceptive processes are a necessary part that enables all these characteristics of our experiences. This is what is distinctive about the Merleau-Pontian view I have developed above: that self and world come into being together through the contribution of interoception to sensing, and exteroception to affect and motor intention. Intero- and exteroception interact and co-enable our experience that stuff exists for us, and our experience of ourselves among that stuff.

Like the quote at the beginning of the chapter suggests, an explication of how we are situated in the present has consequences for how we understand feeling situated in narratives, and in the past and future. All of these things, inhabiting narratives that ostensible are not related to us and our lives, and MTT, are what I am going to call 'inhabiting counterfactuals'. That is, inhabiting times and spaces other that the one we are currently situated in.

It makes sense to suggest that experiences of self in non-occurent situations are not based on totally different processes than those involved in the experiences of self in the present. Rather, a simulation of sensori-affective processes in memory and imagination simulates the feeling of being situated in the present. As Merleau-Ponty puts it, "the body, being our permanent means of "adopting attitudes" and hence of creating pseudo-presents, is the means of our communicating with both time and space." (2012, p. 187). The experience of being situated in counter-factuals is an effect of the body conjuring up a spectacle through the motor intentions and appraisals that come with it. That is, we enter the past and future when we can adopt an orientation, attitude or poise towards something not currently before us[89]. That scene not before us takes shape through these attitudes. We project pseudo-worlds around us as we project our selves. Not only is this parsimonious, but it accurately reflects the evidence that simulations use overlapping systems to online processes, as I discuss shortly.

[89] Kim Atkins (2008) discussion on the interdependence of psychological and bodily continuity overlaps with my discussion here.

Because affect includes its object, and sensory experience includes our sense of self as relating to the world, this interdependence means that as soon as we adopt an attitude towards something beyond the present we summon a scene, and when we imagine a scene, we transport ourselves. These things include each other, so the triggering of one process necessarily implies the other. While a theory of the independence of affect and sensing may explain the same phenomena though claiming these two processes are always, or normally, connected, it has some extra work to do to explain if, and why, this connection co-emerges with perspective. The conjunct of sensed object plus feeling does not obviously explain our sense of perspective. Instead we inhabit counterfactuals through the capacity to integrate intero- and exteroceptive processes.

We have seen that there is always an embodied pre-reflective self in narrative understanding. But what does this mean for moral agency? Our capacity to act on our deepest values?

Well, the idea is that our narrative understanding, constituted through an emotional cadence, is the context through which we understand our current situation. Such an emotional cadence, we now know, co-emerges with a perspective only through integration with sensory experience.

In particular, the integrated sensori-affective character of narrative understanding is the experience of a perspective, and the events it is directed towards, existing through time. That is, we experience ourselves as present now, but within a narrative context that frames our current experience in terms of what led up to it, and the future likely trajectories. The backstory to our current situation and the possible future plots are understood through how they relate to a continuing perspective through time. This is regardless of whether that narrative understanding contains us explicitly or not, because we are prereflectively present, at least as an external perspective, in any act of narrative understanding. Not only does the simulation of intero-exteroceptive processes enable us to experience counterfactuals, but that counterfactual experience is important to our understanding of the present. There is a reciprocity between our imaginative adventures and our capacity to understand the present.

So, for example, if I find myself in a tense exchange with a police officer at a protest, my understanding of the situation, and my choice of action, may be informed by several narratives. First, I probably got into a tense exchange through narratives of the police as agents that maintain the state, including aspects of the state that are violent and oppressive. I don't just see a person in uniform, I see a person who plays a certain role in a narrative about our current society, its formation, how it is maintained, and how it will likely continue, and what all this means for people's lives. Second, my next choice of action will be narratively informed. Through understanding narratives about prison, and the lives of people with criminal records, and stories of people who have been assaulted by police and then accused of assaulting the police, it is likely that I will quickly restrain myself from pursuing the matter further.

All these instances of narrative understanding need not involve mental time travel, if we understand mental time travel to explicitly include the self. As we saw before, my narrative understanding involves a bodily orientation towards the world, the people that populate it, and the events that characterise it. And this can be brought to our understanding of a situation without that bodily orientation explicitly being about our own lives. Neither is metarepresentation necessary for this, as explained in chapter 4.

Further, if embodiment in the present is tied up with our understanding of counterfactuals, through situating our current context within narratives, then we can explain how we can act on our narrative understanding. We have a perspective, which is sensori-affective, on what our current situation is through an understanding of the past and possible future. When our understanding of counterfactuals is narrative, it is always understood in relation to our bodies. Because narrative understanding is also action-orientated, our narrative understanding informs our actions in the present. Our narrative understanding of events consists in a bodily orientation towards those events even while they are not currently happening.

This discussion overlaps with Velleman's *Self to Self* (1996). Like my current discussion, Velleman's interest is how our psychology allows us access to perspectives other than the one we have right now. Velleman suggests that what

matters here is that we are able to construct a scene that "converge[s] at a single point; and it has a *self*-centered scheme of reference because the point of convergence is thought of as occupied by the image's subject" (p. 49-50). For Velleman we have to go further than this though, because here we are visualising, but not yet involved in 'imagined seeing'. In the latter case,

> *The viewer is [present] invisibly, insofar as it now depicts things as seen by him [sic]; and thereby presents him [sic] reflexively, as the subject, in the way that a spoken first-person pronoun presents its speaker.* (p. 51.)

The difference is that, in this case, one does not have to explicitly think of oneself, one is already present 'invisibly'. I disagree with Velleman's distinction here. For me, any scene construction involves an invisible sense of self because sensory experience requires a sense of one's own body. Nonetheless, what I've presented here allows us to understand how such a first-person perspective is enabled: it is through our embodied-embeddedness, an experience that co-emerges with our sensory-affective processing. Velleman's description highlights the form of such a perspective: like a first-person pronoun, it is a reflexive pointing to self, where the point and what is being pointed to converge, and where the pointing happens without explicit thought[90].

But this psychological connection between perspectives also means that we can learn something new by understanding narratives. Particularly, we can learn something new about what actions are available to us in the present, and what they mean. Those actions available to me in narratives are mine in the present because of the identity between the embodied self in the present and in the past and future. This is how, when we imaginatively settle on a best course of action, our reasons for action are reasons on which we can act.

Not only does narrative understanding allow us to imagine scenarios similar to those we've come across before, but it can be creative: we can put together new scenarios. So, the affordances we are faced with now do not just depend on what we have experienced in the past, they also partially depend on the imaginatively created scenarios that we have lived in our heads. And, by virtue of

[90] G&K similarly think that what is important about MTT is that it is an indexical mode of thinking (p. 601).

my interdependence theory, all narrative understanding discloses to us what matters to us, hence our affordances contain a sense of their consequence and value to us. Sensory experience makes narrative understanding possible, and our narrative understanding enhances our general ability to make and act on (moral) decisions.

In sum, if sensory experience and emotional experience are interdependent, then sensory experience in narrative understanding contributes to moral agency in the same way emotions do. Because sensory experiences are involved in emotion, and emotions are necessary for us to have a sense of what matters to us, then sensory experience is involved in this too. Because emotional cadence is necessary for narrative understanding, and narrative understanding is necessary for understanding, and acting on, what we value most highly, sensory experience is necessary for this too.

There are some neuroscientific reasons for thinking that the feeling of being situated when inhabiting counterfactual possibilities is due to the involvement of emotional, motor and sensory processing.

There is reason to think that when we MTT we use emotional, sensory and motor parts of the brain (e.g. Piefke et al. 2003; Markowitsch & Stanilou 2011; Szpunar, Watson & McDermott, 2007). However, the claim here is broader than that, and it is that narrative understanding in general involves sensorimotor and affective processing.

We have seen in the previous chapter that Kaplan et al. (2017) report that the narrative network includes areas that appear to co-ordinate emotional and sensorimotor regions. Further, they took this network to include the medial prefrontal cortex (mPFC), which includes the ventromedial prefrontal cortex (vmPFC).

Sabatinelli et al. (2006) found that brain regions associated with spatial navigation and motor areas were associated with narrative imagery. Xu et al. (2005) in their discussion of narrative processing refer to neuroscientific studies where areas associated with emotion, motor regions, multimodal areas that process exteroceptive information, and areas that become active when asked to

visualise a scene, become active. Similarly, Ferstl et al. (2008) in their meta-analysis found that language, including narrative language, activated the anterior temporal lobes, thought to be involved in memory processing, particular emotional memory, and multimodal processing.

This is evidence that sensory, emotional and motor activation are important for narrative understanding. A good explanation of why this is the case is that they are involved in situating us, in the way that Merleau-Ponty suggests.

5. Conclusion

This chapter has focused on answering a question about the role of sensory experience in moral agency. I have argued that it must be central. Like Prinz, I think that emotions must be pivotal to moral agency, but unlike him, I think that sensory experience is a constituent of our affective sense of what matters, rather than an independent object of that sense. I have presented the view that emotions and sensory experience co-constitute each other, and it is through this intero-exteroceptive process that self and world come into experience, and we can act.

This gives us some explanation of how narrative understanding is possible. When we engage in storytelling we engage a similar intero-exteroceptive process as when we are aware of our current situation. When this sensori-affective process is simulated, it places us in a virtual world. Narrative understanding then becomes central to being able to act for reasons, because it is an embodied process where the actions of our virtual self are available to our present self. Furthermore, as narratively constructed, the meaning of our actions and their consequences are available to us in the present. Thus, like G&K, I take sensory experience to be central to agency.

Of writing a novel, the author Ursula le Guin has said:

> *There is a relationship, a reciprocity between the words and the images, ideas, and emotions evoked by those words: the stronger the relationship, the stronger the work. To believe that you can achieve meaning or feeling without coherent, integrated patterning of the sounds, the rhythms, the*

> *sentence structure, the images, is like believing you can go for a walk without bones*. (1997, p. 196.)

And similarly I have argued imagery, or sensory experience more generally, is a part of the how meaning and feeling are possible during narrative understanding.

Conclusion

"In the beginning was the word." Goethe answered this Biblical phrase through Faust: "In the beginning was the deed." Through this statement, Goethe wished to counteract the word's over-valuation… We can agree with Goethe that the word as such should not be overvalued and can concur in his transformation of the Biblical line to, "In the beginning was the deed." Nonetheless, if we consider the history of development, we can still read this line with a different emphasis: "In the beginning was the deed."
Vygostky, 1987, Thinking and Speech, original emphasis.

I have been focused, in this thesis, on the question: what constitutes our reasons for action? Prinz's answer has been: emotion, and only emotion. My response has been similar to Vygotsky's above: in the *beginning* was emotion. But then there were concepts, and thus the capacity to deliberate. But concepts are affectively constituted, and thus emotion falls within the space of reasons, and enables us to act for reasons.

We started this journey with two theories of moral judgement both with aspects that seemed importantly right. From Prinz, there was something compelling about the idea that emotions are central to moral judgements, and that emotions are embodied and action-orientated. From G&K, we were also introduced to the idea that self-understanding might be crucial to agency, and that subjectively-experienced diachronic thinking, what they called MTT, may be related to self-understanding.

Each theory, however, had their pitfalls and limitations. Prinz is ambivalent about whether moral emotions are conceptual. And, if we look closer at some of the commitments that might underlie G&K's criticism of Prinz, we find McDowellian intuitions that emotions cannot constitute moral judgements unless they are conceptual, in the sense of being able to participate in deliberative reasoning, something that Prinz appears to reject. Reasons for action must be the type of things that can be articulated and defended. Yet G&K's proposal left it puzzling how to relate ideas of MTT and self-understanding with emotions.

So this thesis has attempted to incorporate the insights from both these theories and to resolve some of the issues. I have proposed that acting for reasons is a narrative activity and that it is both rational and experiential. Further, these

processes are mutually forming. Our sensori-affective abilities are constitutive of our capacity to tell and understand stories, but the stories we tell shape our experiences. Specifically, we act for moral reasons when the language we use to tell and understand stories enables us to form what Taylor calls strong evaluations.

Because it is narrative understanding that constitutes the moral sense that we draw on to articulate judgements, moral judgements are not reducible to a singular emotion. Instead they are formed through the relation between the affective concepts that make up the narratives we understand. This theory of moral judgements differs from Prinz's by bringing them into the space of reasons, because our narrative understanding is the type of thing that can be expressed and scrutinized for rational coherence.

In forming this theory, I've rejected some of the commitments of those I have engaged with. I have rejected Prinz's dichotomy between deliberative reasoning and emotions, and the independence of emotion and sensory experience, and argued that these processes are all interdependent. I have rejected G&K's commitment to metarepresentation being central to self-understanding. I have also rejected Velleman's notion that narrative understanding is distinct from metarepresentational understanding, and instead argued that the former is a condition of the latter.

What I want to do now is spell out in a bit more detail how I have managed to respond to Prinz's claim that emotion, and not deliberative reasoning, constitutes moral judgements and arrived at a theory of embodied moral agency. Remember, that this constitution claim is softened by Prinz's more detailed claim that a moral judgement is constituted by an emotion and a representation of the particular object of that emotion. Note, however, that in this proposal it is still the emotion that remains the moral part of the moral judgement, the particular object just anchors the moral sense to its particular object, the particular object does not participate in the moral sense.

How does Prinz manage to argue for this conclusion, that emotions constitute our moral attitudes? One thing he has to do, to defend this, is also argue for the independence of various psychological processes. Included in this project,

is establishing that our explicit rational capacities can only be causally relevant to our emotions, rather than constituents of them. If emotions and deliberative reasoning are interconnected in some constitutive way, then the possibility of Prinz's moral emotions, distinct from deliberative capacities, constituting judgements falls away.

In my first chapter, I argued that Prinz equivocated on whether moral emotions constitute concepts or not. Although he stated that they do, he does not appear to think the moral emotions are part of our deliberation, instead they are a result of it. However, concepts, for Prinz, are normally things that can participate in deliberation. It seems that he is ambiguous on this point because he is trying to defend sentimentalism. As such, making deliberation constitutive of moral judgements would by a synthesis of rationalism and sentimentalism, rather than a clearly sentimentalist position.

In my second chapter, I introduced, using McDowell, an alternative way to understand 'judgement'. Here, for something to be a judgement, it is the articulation of our conceptual, but implicit, reasons. Further, for it to count as a judgement, it must be justified and not merely caused. Prinz, as understanding moral judgements as merited because they are caused in the right way, rather than justified, fails to give a theory of moral judgements. My argument, however, requires that the language we use to characterise rational agents is distinct and legitimate. Meanwhile causal explanation can only explain the enabling conditions that make creatures like ourselves possible. In doing this, I make a fundamental commitment to how to best understand questions of agency and knowledge.

In chapter 3, I started developing my own theory of what type of things we are as agents. Here, I looked at G&K claim that MTT is essential for us to be able to make and act on responsible decisions. I argued that G&K are right that being able to inhabit a sequence of events makes a special contribution to making and acting on decisions, but that they are wrong that this is specific to MTT. Instead, I argued that the general capacity to inhabit counterfactuals, called narrative understanding – rather than the more particular capacity to inhabit the past and future of what is explicitly depicted as our own life – is important for decision-making. Further, I proposed that if we understand emotions as embodied appraisals, as Prinz does,

then we can explain the phenomenology of narrative understanding and MTT, and why that phenomenology co-occurs with the possibility of action. As Velleman explains, narrative understanding is characterised by an emotional cadence. So, emotions, as providing us with a felt understanding of how we stand in relation to a situation, co-emerge with a sense that we are inhabiting a situation.

In chapter 4, I related narrative understanding to the theory that we are agents when we can act in light of our own self-understanding. I argued that narrative understanding is always implicitly self-understanding because understanding a narrative always contains a pre-reflective experience of our selves and how we relate to a situation. Because of this, the ability for narrative understanding to unify our understanding of the world is also the potential to unify our implicit understanding of our selves. If reasons for actions consist of acting in light of what would make sense for us to do, given our commitments, then narrative understanding enables us to act on reason. Unlike Velleman, I argued that narrative understanding and metarepresentational understanding are on a continuum – our ability for narrative understanding emerges with a diachronic, teleological, perspective. The implicit understanding of such a perspective is necessary to explicitly think about our selves and others as thinking things that pursue goals over time.

In my fifth chapter, I considered how narrative understanding is involved in moral agency. It is in this chapter that I returned to Prinz and the relationship between emotions and concepts, and I gave further reasons that supported the alternative understanding of 'judgement' that I had started in chapter 2. Judgements are the act of a language-using creature because, by enabling us to stand back and reflect on our attitudes, language enables us to be responsible for our own perspective. Looking at Taylor's theory of strong evaluations, I explained how the moral sense through which we make moral judgements emerges with narrative understanding that involves moral language. Moral narrative understanding provides us with a moral sense of what is of qualitatively higher and lower worth through providing us with a coherent network of affective reasons.

This theory of moral agency is supported by a phenomenological description of our moral thinking as one that is continually both affective and conceptual, and empirical evidence that shows concepts, particularly abstract ones, are affective. Both these pieces of evidence fit with the theory that emotions, through being constitutive of concepts, can be part of the space of reasons. Since Prinz understands emotions as nonconceptual in the sense that they are not involved in deliberation, this also departs from Prinz by arguing that it is not a singular emotion joined to representation of their particular object that constitutes a judgement, but articulation of the interrelation of various affective concepts. As an alternative to Prinz, therefore, the theory of embodied narrative moral agency gives a detailed account of what psychological processes enable us to act for moral reasons, and allows us to make sense of phenomenological and empirical considerations. While it does not disprove Prinz's theory, it does provide a competing explanation that doesn't result in the highly counter-intuitive claim that only emotions, and not deliberative processes, are constitutive of moral judgements.

In my final chapter, I argued that the theory that emotions and sensory experience are interdependent also provides us with an account of how narrative understanding is possible, and how we can act on it, and has theoretic virtues that Prinz's theory of the independence of sensory and emotional experience lacks. I draw on the phenomenology of Merleau-Ponty to explain how our sense of being present in the present consists of the world being experienced affectively, and our bodies being experienced as engaged with the situation we are in. This can be explained neurologically, through the intertwining of intero- and exteroceptive processes in perception. Unsurprisingly, then, when there is an intense and global break down in the co-ordination of intero-exteroceptive processing, such as in psychosis, people experience an altered sense of self and world. That is, the interdependence of sensory and emotional experience can explain cases in clinical psychology. Finally, this contributes to an explanation of how we can act on our narrative understanding through explaining how it is possible to have a perspective on our current situation. My proposal is that our pre-reflective sense of embodiment in our narrative imaginings is a simulation of the same intero-

exteroceptive processes that provide us with our current sense of embodiment. This makes narrative understanding an indexical type of understanding. Both our actual and virtual embodiment prereflectively point back to the same body. Because our embodiment is also thoroughly action-orientated, narrative understanding discloses possible actions to us, and so the possibilities for action we encounter through narrative understanding are available to us now.

In the end, this thesis expands on a truism about what type of things we are as humans, as agents. So we've gone on this journey to arrive near the beginning of the western philosophical tradition. Aristotle expressed it long ago: we are rational animals. Here, I have suggested that our rationality is in virtue of our embodiedness and our emotionality, in virtue of us having a world that matters to us. Prinz is right to stress the importance of emotion to being agents, he is right that because we have affect, we can act, and we can act on what matters to us. He is right that this relation is one of constitution. But G&K are right too to stress the importance of self-understanding, and the importance of rational thought to moral agency.

This led us to the importance of narrative understanding for moral agency. Narrative understanding is the capacity to inhabit a sequence of events; it allows us to be present in virtual situations and for virtual situations to contextualise our current predicament. When we can produce narratives, we have agency, we have the capacity to act on what makes experiential and logical sense to us. A capacity that is conceptual-affective-sensory. Because these processes cannot be pulled apart fully it does not make sense to say that emotion alone constitutes agency because it is mutually formed with our capacity for concepts and sensory experience. I've suggested that this is the best way to interpret the evidence from our experiences and from science, a way that fits with a philosophical account of agency, where we are agents because we act for reasons. Reasons, now, are constituted through an emotional cadence, but this is part of, and not counter to, our deliberative capacities.

References

Ahmed, S. (2013). *The cultural politics of emotion*. Routledge.

Atkins, K. (2008). Narrative identity and embodied continuity. In *Practical identity and narrative agency*. (Ed: Atkins, K., & Mackenzie, C.) Routledge: 78-98.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states?. *Psychological review*, *116*(4), 953.

Barbas, H., Saha, S., Rempel-Clower, N., & Ghashghaei, T. (2003). Serial pathways from primate prefrontal cortex to autonomic areas may influence emotional expression. *BMC neuroscience*, *4*(1), 1.

Barlassina, L., & Newen, A. (2014). The role of bodily perception in emotion: In defense of an impure somatic theory. *Philosophy and Phenomenological Research*, *89*(3), 637-678.

Barrett, L. F., & Bar, M. (2009). See it with feeling: affective predictions during object perception. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1521), 1325-1334.

Bermúdez, J. L. (2001). Nonconceptual self-consciousness and cognitive science. *Synthese*, *129*(1), 129-149.

Bloch, M. (2011). The blob. *Anthropology of this Century*, 1.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*(3), 242-261.

Borghi, A. M., & Binkofski, F. (2014). Language, languages, and abstract concepts. In *Words as Social Tools: An Embodied View on Abstract Concepts*. Springer New York, 111-124.

Bratman, M. E. (2000). Reflection, planning, and temporally extended agency. *The Philosophical Review*, *109*(1), 35-61.

Bylsma, L. M., Morris, B. H., & Rottenberg, J. (2008). A meta-analysis of emotional reactivity in major depressive disorder. *Clinical psychology review*, *28*(4), 676-691.

Carruthers, P. (2009). An architecture for dual reasoning. In *In Two Minds: Dual Processes and Beyond* (Ed: Evans, J & Frankish, K.). Oxford University Press, 109-27.

Claridge, G., McCreery, C., Mason, O., Bentall, R., Boyle, G., Slade, P., & Popplewell, D. (1996). The factor structure of 'schizotypal 'traits: a large replication study. *British Journal of Clinical Psychology*, *35*(1), 103-115.

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Cole, J. (1995). *Pride and a daily marathon*. MIT Press.

Colombetti, G. (2014). *The feeling body: Affective science meets the enactive mind.* MIT press.

Colombetti, G. (2007). Enactive appraisal. *Phenomenology and the Cognitive Sciences*, *6*(4), 527-546.

Craver, C. F., Kwan, D., Steindam, C., & Rosenbaum, R. S. (2014). Individuals with episodic amnesia are not stuck in time. *Neuropsychologia*, *57*, 191-195.

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. Houghton Mifflin Harcourt.

de Haan, S., & Fuchs, T. (2010). The ghost in the machine: disembodiment in schizophrenia–two case studies. *Psychopathology*, *43*(5), 327-333.

DeLancey, C. (2002). *Passionate engines: What emotions reveal about the mind and artificial intelligence*. Oxford University Press.

Döring, S. A. (2007). Seeing what to do: Affective perception and rational motivation. *Dialectica*, *61*(3), 363-394.

Dunn, R. (1998). Knowing what I'm about to do without evidence. *International journal of philosophical studies*, *6*(2), 231-252.

Elgin, C. Z. (2008). Emotion and understanding. In *Epistemology and Emotions*. Ashgate. 33-50.

Feinberg, M., Willer, R., Antonenko, O., & John, O. P. (2012). Liberating reason from the passions: Overriding intuitionist moral judgments through emotion reappraisal. *Psychological science*, *23*(7), 788-795.

Ferstl, E. C., Neumann, J., Bogler, C., & Von Cramon, D. Y. (2008). The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Human brain mapping*, *29*(5), 581-593.

Franklin, C. E. (2015). Self-determination, self-transformation, and the case of Jean Valjean: a problem for Velleman. *Philosophical Studies*, *172*(10), 2591-2598.

Frijda, N. H. (2004). Emotions and action. In *Feelings and emotions: The Amsterdam symposium* (Ed: Manstead, A., Frijda, N.H., & Fischer, A). Cambridge University Press, 158-173.

Gardner, S. (2015). Merleau-Ponty's transcendental theory of perception. In *The Transcendental Turn* (Ed: Gardner, S. & Grist, M.). Oxford University Press.

Gerrans, P., & Kennett, J. (2010). Neurosentimentalism and moral agency. *Mind*, 119 (475): 585-614.

Gerrans, P., & Kennett, J. (2017). Mental time travel, dynamic evaluation, and moral agency. *Mind:* 126 (501): 259-268.

Gigerenzer, G. (2004). Fast and frugal heuristics: The tools of bounded rationality. *Blackwell handbook of judgment and decision making*, 62-88.

Gilboa, A. (2004). Autobiographical and episodic memory—one and the same?: Evidence from prefrontal activation in neuroimaging studies. *Neuropsychologia*, *42*(10), 1336-1349.

Goldie, P. (2012). *The mess inside: narrative, emotion, and the mind*. Oxford University Press.

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in cognitive sciences*, *11*(7), 299-306.

Heller, A. S., Johnstone, T., Shackman, A. J., Light, S. N., Peterson, M. J., Kolden, G. G., & Davidson, R. J. (2009). Reduced capacity to sustain positive emotion in major depression reflects diminished maintenance of fronto-striatal brain activation. *Proceedings of the National Academy of Sciences*, *106*(52), 22445-22450.

Hume, D. (2006). *An enquiry concerning the principles of morals* (Vol. 4). Oxford University Press.

Hurley, S. (2008). The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behavioral and Brain Sciences*, *31*(01), 1-22.

Hurley, S. (2006). Making sense of animals. *Rational animals?* (Ed: Hurley, S., & Nudds, M.). Oxford University Press, 139-171.

Hurley, S. (2003). Animal action in the space of reasons. *Mind & Language*, *18*(3), 231-257.

Hurley, S. L. (1997). Nonconceptual self-consciousness and agency: Perspective and access. *Communication and Cognition (Part 1 of Special Issue: Approaching Consciousness)*, *30*(3/4), 207-248.

Hurley, S., & Noë, A. (2003). Neural plasticity and consciousness. *Biology and Philosophy*, *18*(1), 131-168.

Hutto, D. D. (2012). Truly enactive emotion. *Emotion Review*, *4*(2), 176-181.

Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly*, *32*(127), 127-136.

James, W. (1884). II. What is an emotion?. *Mind*, (34), 188-205.

Kaplan, J. T., Gimbel, S. I., Dehghani, M., Immordino-Yang, M. H., Sagae, K., Wong, J. D., & Damasio, A. (2017). Processing Narratives Concerning Protected Values: A Cross-Cultural Investigation of Neural Correlates. *Cerebral Cortex*, *27*(2), 1428-1438.

Kawabata, H., & Zeki, S. (2004). Neural correlates of beauty. *Journal of neurophysiology*, *91*(4), 1699-1705.

Korsgaard, C. (2006). Morality and the distinctiveness of human action. In *Primates and philosophers: How morality evolved*. (Ed: F., Macedo, S., & Ober, J.) Princeton University Press, 98-119.

Korsgaard, C. M. (1989). Personal identity and the unity of agency: A Kantian response to Parfit. *Philosophy & Public Affairs*, 18(2), 101-132.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14.

Kwan, D., Craver, C. F., Green, L., Myerson, J., Boyer, P., & Rosenbaum, R. S. (2012). Future decision-making without episodic mental time travel. *Hippocampus*, *22*(6), 1215-1219.

Lambert, M. V., Senior, C., Fewtrell, W. D., Phillips, M. L., & David, A. S. (2001). Primary and secondary depersonalisation disorder: a psychometric study. *Journal of affective disorders*, *63*(1), 249-256.

Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American psychologist*, *46*(8), 819.

LeDoux, J. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life.* New York: Touchstone.

Le Guin, U. K. (1997). *Dancing at the edge of the world: Thoughts on words, women, places*. Grove Press.

Levenson, R. W., Ekman, P., & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, *27*(4), 363-384.

Levine, B., Turner, G. R., Tisserand, D., Hevenor, S. J., Graham, S. J., & McIntosh, A. R. (2004). The functional neuroanatomy of episodic and semantic autobiographical remembering: a prospective functional MRI study. *Journal of Cognitive Neuroscience*, *16*(9), 1633-1646.

Levy, N. (2014). Consciousness, implicit attitudes and moral responsibility. *Noûs*, *48*(1), 21-40.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and Brain Sciences*, *35*(03), 121-143.

Mackenzie, C. (2007). Bare personhood? Velleman on selfhood. *Philosophical Explorations*, *10*(3), 263-281.

Markowitsch, H. J., & Staniloiu, A. (2011). Memory, autonoetic consciousness, and the self. *Consciousness and cognition*, *20*(1), 16-39.

Mar, R. A. (2004). The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia*, *42*(10), 1414-1434.

Mason, O., Claridge, G., & Jackson, M. (1995). New scales for the assessment of schizotypy. *Personality and Individual differences*, *18*(1), 7-13.

McDowell, J. (2009). Conceptual capacities in perception. *Having the world in view: essays on Kant, Hegel, and Sellars*. Harvard University Press, 127-44.

McDowell, J. (2008). Responses. *John McDowell: Experience, norm, and nature*. (Ed: Lindgaard, J.) Blackwell Publishing, 200-267.

McDowell, J. (1994). *Mind and world*. Harvard University Press.

Meijsing, M. (2000). Self-consciousness and the body. *Journal of consciousness studies*, *7*(6), 34-52.

Metzinger, T. (2004). *Being no one: The self-model theory of subjectivity*. MIT Press.

Merleau-Ponty, M. (2012). *Phenomenology of perception*. Routledge.

Milgram, S. (1974). *Obedience to authority: An experimental view.* Harpercollins

Milgram, S. (1963). Behavioral study of obedience. *The Journal of abnormal and social psychology*, *67*(4), 371.

Morris, J. S., Öhman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proceedings of the National Academy of Sciences*, *96*(4), 1680-1685.

Morrison, T. (2004, May 28). *College Commencement Speech*. Speech presented in Wellesley College. Retrieved October 14, 2016, from https://www.youtube.com/watch?v=SAJH03U7aHM

Moskowitz, G. B., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *Journal of Experimental Social Psychology*, *47*(1), 103-116.

Nagel, T. (1974). What is it like to be a bat?. *The philosophical review*, *83*(4), 435-450.

Nietzsche, F. (1961). Thus spoke Zarathustra, trans. *RJ Hollingdale*

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature reviews neuroscience*, *9*(2), 148-158.

Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., & Goodyear, B. G. (2007). Neural correlates of concreteness in semantic categorization. *Journal of Cognitive Neuroscience*, *19*(8), 1407-1419.

Piefke, M., Weiss, P. H., Zilles, K., Markowitsch, H. J., & Fink, G. R. (2003). Differential remoteness and emotional tone modulate the neural correlates of autobiographical memory. *Brain*, *126*(3), 650-668.

Phillips, L. K., & Seidman, L. J. (2008). Emotion processing in persons at risk for schizophrenia. *Schizophrenia Bulletin*, *34*(5), 888-903.

Prinz, J. (2007). *The emotional construction of morals*. Oxford University Press.

Prinz, J. (2006). The emotional basis of moral judgments. *Philosophical explorations*, *9*(1), 29-43.

Prinz, J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford University Press.

Ratcliffe, M. (2013). Depression and the phenomenology of free will. In *The oxford handbook of philosophy and psychiatry*. (Ed: Fulford, K., Davies, M., Gipps, R., Graham, G., Sadler, J., Stanghellini, G., & Thorton, T.) Oxford University Press, 574-591.

Ratcliffe, M. (2005). The feeling of being. *Journal of Consciousness Studies*, *12*(8-9), 43-60.

Roskies, A. (2003). Are ethical judgments intrinsically motivational? Lessons from "acquired sociopathy'. *Philosophical Psychology*, *16*(1), 51-66.

Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in cognitive sciences*, *16*(3), 147-156.

Rukeyser, M. (1968). *The speed of darkness.* Random House.

Sabatinelli, D., Lang, P. J., Bradley, M. M., & Flaisch, T. (2006). The neural basis of narrative imagery: emotion and action. *Progress in brain research*, *156*, 93-103.

Saver, J. L., & Damasio, A. R. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia*, *29*(12), 1241-1249.

Sellars, W. (1963). Philosophy and the scientific image of man. *Science, perception and reality*, *2*, 35-78.

Sellars, W. (1956). Empiricism and the Philosophy of Mind. *Minnesota studies in the philosophy of science*, *1*(19), 253-329.

Schechtman, M. (2007). Stories, lives, and basic survival: A refinement and defense of the narrative view. *Royal Institute of Philosophy Supplement*, *60*, 155-178.

Slaby, J., & Stephan, A. (2008). Affective intentionality and self-consciousness. *Consciousness and Cognition*, *17*(2), 506-513.

Slaby, J. (2012). Affective self-construal and the sense of ability. *Emotion Review*, *4*(2), 151-156.

Smith, M. (1996). The argument for internalism: Reply to Miller. *Analysis*, *56*(3), 175-184.

Smith, M. R. (1994). *The moral problem.* Blackwell.

Strawson, G. (2004). Against narrativity. *Ratio*, *17*(4), 428-452.

Subotnik, K. L., & Nuechterlein, K. H. (1988). Prodromal signs and symptoms of schizophrenic relapse. *Journal of Abnormal Psychology*, *97*(4), 405.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans?. *Behavioral and Brain Sciences*, *30*(03), 299-313.

Szpunar, K. K., Watson, J. M., & McDermott, K. B. (2007). Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences*, *104*(2), 642-647.

Tan, T. E. (2012). *The Garden of Evening Mists: A Novel*. Weinstein Books.

Taylor, C. (2007). A Secular Age. Cambridge.

Taylor, C. (1985). *Human agency and language.* Cambridge University Press.

Taylor, C. (1983). Hegel's philosophy of mind. In *Philosophy of Mind/Philosophie de l'esprit* (Ed: Reinstad, D.). Springer Netherlands, 133-155.

The Fox Without a Tail An Aesop's Fable. Retrieved October 05, 2016, from The Fox Without a Tail an Aesop's Fable By Alchin - http://www.taleswithmorals.com/aesop-fable-the-fox-without-a-tail.htm

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.

Vartanian, O., & Goel, V. (2004). Neuroanatomical correlates of aesthetic preference for paintings. *Neuroreport*, *15*(5), 893-897.

Vargas, M. R. (2013). Situationism and moral responsibility: free will in fragments. In *Decomposing the Will* (Ed: Clark, A., Kiverstein, J. & Vierkant, T.) Oxford University Press, 325-350.

Velleman, J. D. (2007). Reply to Catriona Mackenzie. *Philosophical Explorations*, *10*(3), 283-290.

Velleman, J. D. (2006). *Self to self: Selected essays*. Cambridge University Press.

Velleman, J. D. (2003). Narrative explanation. *The philosophical review*, *112*(1), 1-25.

Velleman, J. D. (2000). From self psychology to moral philosophy. *Noûs*, *34*(s14), 349-377.

Velleman, J. D. (1996). Self to self. *The Philosophical Review*, *105*(1), 39-76.

Velleman, J. D. (1992). What happens when someone acts?. *Mind*, *101*(403), 461-481.

Vigliocco, G., Kousta, S. T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: the role of emotion. *Cerebral Cortex*, *24*(7), 1767-1777.

Vodušek, V. V., Parnas, J., Tomori, M., & Škodlar, B. (2014). The phenomenology of emotion experience in first-episode psychosis. *Psychopathology*, *47*(4), 252-260.

Vygotsky, L. S. (1934). Thinking and speaking. Retrieved July 1, 2016, from Marxists Internet Archive - https://www.marxists.org/archive/vygotsky/works/words/ch07.htm

Walter, H. (2012). Social cognitive neuroscience of empathy: concepts, circuits, and genes. *Emotion Review*, *4*(1), 9-17.

Wilson-Mendenhall, C. D., Simmons, W. K., Martin, A., & Barsalou, L. W. (2013). Contextual processing of abstract concepts reveals neural representations of nonlinguistic semantic content. *Journal of Cognitive Neuroscience*, *25*(6), 920-935.

Wheeler, M. (2013). Science friction: Phenomenology, naturalism and cognitive science. *Royal Institute of Philosophy Supplement*, *72*, 135-167.

Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: emergent features of word, sentence, and narrative comprehension. *Neuroimage*, *25*(3), 1002-1015.

Yung, A. R., Phillips, L. J., Yuen, H. P., Francey, S. M., McFarlane, C. A., Hallgren, M., & McGorry, P. D. (2003). Psychosis prediction: 12-month follow up of a high-risk ("prodromal") group. *Schizophrenia research*, *60*(1), 21-32.

Yung, A. R., & McGorry, P. D. (1996). The prodromal phase of first-episode psychosis: past and current conceptualizations. *Schizophrenia bulletin*, *22*(2), 353-370.