# *Introduction*

## 0.1 The Big Picture

Once we start wondering about our emotions, it's hard to stop. Their pursuit leads the unwary traveller down a rabbit hole with tunnels connecting to every part of the mind. For the average emotional experience contains much more than curious sensations. It is liable to involve quite complex evaluations, pleasures and pains, associations, imaginings, the awareness of bodily capability, sensitivity to social norms, directed reasonings, action plans, attempts at self-regulation and long-term commitments, besides the influence of individual personality traits. And all this jostles alongside the subject's other concerns and projects, each bearing emotional implications of their own.

To grasp this complexity is the principal aim of this book. But my strategy is not that of a novelist offering nuanced depictions of lived experience. I aim to proceed systematically – to divide the mind up into relatively discrete levels and proceed from bottom to the top, pointing out the key components and structural relationships. At the foundations, I want to introduce a new sort of mental state: valent representation. Valent representation is built around the basic principle of negative feedback control.[1] Its function is to detect, with varying degrees of intensity, the presence or absence of specified conditions. The valent (i.e. positive or negative) part is that, without needing the mediation of any other representational state, a response is triggered that is disposed to either increase or decrease the presence of the condition being

[1] Negative feedback loops form the basis of 'cybernetics' or control theory, which generated widespread discussion following the publication of Norbert Wiener's influential 1948 work. There was a time when 'cybernetic' was one of those hip academic buzzwords like 'paradigm' or 'non-linear' and became so overused that it lost all meaning. The term hasn't been fashionable for a while now, though control systems continue to be central to the study of both artificial and biological systems. Rodney Brooks (e.g. 1991) in particular is known for applying control systems to robotic simulations of cognition.

represented. In this way, valent representations regulate the presence of the conditions they track.

Another important feature of valent representation is that it is generative; its components can be elaborated to provide more sophisticated forms of control. It is by means of such elaborations that I will develop models of the various affective states, including the emotions. Moreover, each new elaboration builds upon earlier stages of control. So by progressing through the affective states we develop an architecture, where control builds on control, until we end up with a general model of the mind. At the end of this book I will present a sketch of this general architecture.

I should stress then, that while proceeding systematically, I am offering a big picture account of the emotions and related affective phenomena, including the way they show up in sophisticated cognitive activity. However, to present this picture in a relatively encapsulated form, I must move fairly swiftly (by academic standards) from one area of our affective lives to another. I hope the reader will forgive me if there are times when I push on rather than worry about every possible alternative account. No doubt, there will also be moments where some readers will want to pause and reflect on the finer details. However, my first priority is to convey a sense of how it all fits together.

## 0.2 In Search of Synthesis

To motivate this approach to emotions and related states, it is helpful to begin with a brief review of the current state of emotion research. As I see it, the contemporary state of emotion theory is characterised by the search for a synthesis between two general approaches: On the one hand we have the cognitivist position, drawing upon a tradition going back as far as Aristotle and the Stoics, according to which emotions are a species of evaluative judgement. On the other, we have the noncognitivist or somaticist position, according to which emotions are essentially patterns of bodily feelings. This latter position is most identified with William James, particularly his seminal article 'What is an emotion?' (1884).[2]

[2] James' approach was subject to severe criticism in the early part of the twentieth century, and the cognitivist view became dominant. A notable debate between the two positions is Zajonc (1984) and Lazarus (1984) (see Prinz 2004: chapter 2 for a detailed commentary). While it is fair to say that the cognitivist view still dominates in psychology, the situation is less clear in philosophy. The Jamesian view enjoyed a resurgence in the 1990s, in part due to the work of Antonio Damasio (1994). A recent philosophical proponent is Jenefer Robinson (1995; 2005).

Both theories definitely have something going for them. Unfortunately, they are hard to combine elegantly.

My take on emotions has always started with the claim that bodily responses play a vital role in emotional states. I am initially drawn by James' thought experiment in which he asks us to imaginatively subtract the feelings of bodily activity accompanying an emotion. It seems that when we subtract the various bodily sensations of grief, for example, all we are left with is the 'feelingless cognition that certain circumstances are deplorable', not an emotion. I personally find this thought experiment compelling, but others seem less impressed. In Chapter 4 I will offer some reasons for why this might be so.

Apart from how (some? most?) people think or talk about emotions, it is clear that there must be at least a strong reciprocal causal relationship between emotions and bodily responses. Emotional episodes are observed to stimulate various behaviours such as fighting or fleeing, as well autonomic bodily responses like raised heart rate or the release of adrenaline. Meanwhile, manipulating one's bodily state, either by means of expressive behaviours or drugs, can alter both observed and self-reported emotions (see e.g. Laird 2007 for a review). How fixed the relationship is between particular emotional states and patterns of bodily responses is a matter of some dispute (for discussion see Section 3.7). However the general connection between emotions and bodily responses is indisputable.

While there are clear causal connections between emotions and bodily responses, this falls short of the claim that emotions are constituted by bodily responses. We can get a bit closer to such a claim by emphasising the probable evolutionary provenance of emotions. That is, we observe comparable bodily responses and expressive behaviours in many other species. This suggests that emotions are a faculty of ancient evolutionary pedigree. In addition, responses like fleeing in fear or crying for help confer obvious survival benefits. Combined, these points make it reasonable to suppose that bodily responses are a key evolved function of emotion; were it not for the triggering of bodily responses, emotions would not have been selected or preserved by evolutionary processes. To the extent that we allow evolved functionality to define psychological states, we can then assert that, whatever else they may be, emotions are *dispositions* or *tendencies* to respond bodily. This is the position of psychologist Nico Frijda for one

> Emotions, then, can be defined as modes of relational action readiness, either in the form of tendencies to establish, maintain, or disrupt a

> relationship with the environment or in the form of mode of relational readiness as such. (Frijda 1986: 71)[3]

Note that if emotions are dispositions to respond, this allows for the possibility that a response may be suppressed or interrupted before overt behaviour is triggered. There are several stages of bodily preparation that must occur prior to a response like running away. The sympathetic nervous system of the body must first be aroused in various ways, raising our heart and breathing rates. And before this, the brain will need to issue various motor instructions. Given this causal structure, it is hard to identify a precise stage where we can say 'now an emotion exists'. But we can at least be sure that an emotion has been manifested once a physiological change is triggered.

Meanwhile, Frijda thinks that what he calls 'modes of action readiness' must be stimulated or guided by cognitive appraisals of the situation. So let us turn to the claims of cognitivism. In the recent philosophical era, cognitivism has been most closely associated with the works of Robert Solomon (e.g. 2007) and Martha Nussbaum (e.g. 2003). It is also closely allied with appraisal theories in psychology (e.g. Lazarus 1991; Scherer 2005). The key point made by the cognitivists is that we are not happy or sad simpliciter, but happy or sad *about* something. That is, emotions are intentional or content bearing mental states. Additionally, emotions seem to aim at the factual state of affairs in a manner comparable to beliefs and judgements. For instance, we are sad when something bad has in fact happened or afraid when something really is threatening. Even an emotion like hope may rely on a judgement that a certain state of affairs is in fact available, given what is currently known to be the case (cf. McCormick 2016).

Supporting this point about emotional meaning is the observation that we hold up our emotions to standards of rational evaluation. We blame someone for feeling angry about a misconstrued offense, or a phobic fear of something harmless. We may even blame someone for feeling hope towards an outcome that we know to be impossible. Thus the absence of the relevant facts undermines the fittingness of the emotional state. It is also worth emphasising that emotions bear rational connections with each other. If you are scared that your precious vase (or nose) might get broken, you are thereby committed to being sad if it is in fact broken, and relieved

[3] In his last publication, Frijda (2016) also suggested that affective states are built upon negative feedback; however, the view is not developed very carefully.

if it isn't (e.g. Helm 2001: 152–156). This implies that emotions have semantic content holding them up to norms of rational consistency.

The greatest strength of cognitivism is the greatest weakness of the somaticist view. How can a set of bodily responses be about the state of the world outside of one's body? Yet the somaticist may reply, how can an emotion-bearing creature like a cat or a prelinguistic infant make judgements of the sort the cognitivist demands? And so begins a very involved debate. It is not my aim in this book to provide yet another survey of its various ins and outs. More patient introductions can be found in Deonna and Teroni (2012) and Price (2015).[4] I will proceed on the assumption that the previous points are agreed by nearly everyone working in emotions theory.

So far, it seems we must find some way to link the semantics of emotions with the disposition to respond in certain ways. Moreover, this link must operate at a pretty basic cognitive level, available to nonlinguistic animals and infants (though certain 'higher' emotions like guilt and shame might be reserved for cognitively sophisticated creatures). Of course we could just say that emotions are capacities that mix a representational component with a reactive component. However this is a deeply unsatisfying answer from the point of view of biological plausibility, because if both intentional content and dispositions to respond are necessary features of emotion, then this is unlikely to be a mere matter of coincidence. On the contrary, we should expect that the key adaptive innovation of emotions lies in how these two features are functionally connected.

The most obvious way to specify the connection is that the representation guides the bodily reaction. However the somaticist is likely to complain that such an approach makes the appraisal the real business end of emotions and the response merely its causal product. They insist that the state would not be an emotion at all without the bodily response. Indeed, Jamesians have wanted to reverse the priority and claim that the bodily reaction comes first, and the distinctive feature of emotional intentionality is in noticing that one is disposed to react in a certain way.

Thus it is never sufficient to simply say that emotions combine representations and reactions. If we are to properly understand emotions, we must understand the precise nature or manner of this combination.[5]

[4] Other compelling critiques of the cognitivist approach are Robinson (1995) and Scarantino (2010).
[5] This is the main reason why I find a recent model proposed by Barlassina and Newen (2013) unsatisfying. These philosophers argue that in emotional states, perceptions of bodily responses and representations of external objects are 'integrated' (somehow) by the subject. But I want to know exactly how this integration is achieved.

## 0.3 The Perceptual Theory

One of the most significant proposals in recent years for synthesising the cognitive and somatic aspects of emotions is the perceptual theory of emotion. There are precedents for this theory in the work of Ronald de Sousa (1987); however, it is most closely associated with Jesse Prinz (2004) as well as Christine Tappolet (2000; 2016) and Sabine Döring (2007). Prinz synthesises the cognitive and somaticist positions by claiming that the intentional, world-directed content of emotions is achieved by means of bodily activity. Specifically, he claims that evolution has set up creatures to reliably trigger certain patterns of bodily responses whenever situations displaying certain qualities are detected (e.g. danger, offence, loss). The individual's feeling of this pattern of bodily responses then is the perception of the situational quality. In effect, Prinz claims, patterns of bodily changes are serving a comparable function to patterns of retinal activity; they reveal to the subject the presence of dangers, offences, losses and so on.

The perceptual theory has some advantages over cognitivism. First, if emotions are perceptions rather than judgements, this can explain what is known as 'emotional recalcitrance'; the fact that emotions may persist despite judgements to the contrary. For instance, we can remain susceptible to an emotion like fear, or have a lingering sense of anger, despite explicitly judging there to be no harm or intended offense. The perceptual theorist argues that while our evaluative judgements are not normally isolated from our wider belief states in this way, perceptual experience can be. Things can perceptually seem a certain way despite our better judgements. The standard example is the Müller-Lyer illusion, in which the two lines appear different lengths despite the belief that they are equal. Thus, recalcitrant emotions can now be interpreted as a kind of illusory experience. Both emotions and perceptions aim at the factual condition of the world, but both can get it wrong in ways that are potentially independent of our beliefs (see in particular Döring 2007 for discussion).

Several other comparisons between emotional experience and perceptual experience can be drawn. Both are episodic, both have a qualitative feel, and both are relative to the perspective of the individual. Prinz even claims that there are physiological analogies between emotions and perceptions, since although we don't have an emotion organ on the surface of our bodies, emotions may have dedicated processing modules (the amygdala, an almond-shaped structure near the centre of the brain, is typically implicated here) (Prinz 2004: chapter 10).

Of course, there is at least one crucial disanalogy between emotions and perceptions, which is that emotions are evaluative in nature whereas perceptions are purely descriptive. However, this may be dealt with by merely specifying that the presentation of values is the key distinctive feature of emotion perceptions as opposed to other sorts of perceptions (see especially Tappolet 2000; 2016).

The perceptual theory seems to improve upon the cognitive theory in giving us a low-level way to capture information. Yet the perceptual theory has rather fallen out of favour in the last few years. Various attacks can be found in Salmela (2011), Deonna and Teroni (2012: chapter 6), Barlassina and Newen (2013) and Brady (2013). One important objection concerns the relationship between emotions and reasons. I mentioned earlier that we identify reasons for feeling the way we do, and we often demand reason-based explanations for emotions, particularly when their appropriateness is challenged. As Michael Brady in particular has been keen to point out, this is not the case for perceptual states. Perceptual states are not reason responsive. Thus there is a significant epistemic disanalogy.

A related worry about perceptual theory is that the arousal of emotions relies completely upon our other perceptual and cognitive capacities. Emotions are only triggered because we are first acquainted with situations by other means. In particular, emotions *routinely* rely on our imaginings, memories and empathy for others (cf. Morton 2013). Meanwhile, bodily feelings are themselves drawn from our interoceptive faculties for detecting visceral changes, blood pressure, balance, muscle tension and so on.[6] This implies that emotions are not a means by which the organism draws information into the system, as we find with perception. Rather emotions rely on performing some kind of operation on the information we draw in, as we find with cognition.[7]

A further problem with the perceptual theory concerns phenomenology. Prinz draws an analogy between perceiving dangers in virtue of feeling bodily changes and seeing colours in virtue of retinal activity. Yet unlike the case of retinal activity, our bodily changes are not phenomenally transparent. That is, bodily changes are not something we fail to notice

[6] Cf. Hatzimoysis who in 2003 argued that perception is not transitive such that by perceiving one thing (i.e. bodily changes by means of interoception) you can then also perceive its causal trigger (the offence). For example, by perceiving a cloud you don't perceive the evaporation processes that built it. For some reason this argument doesn't seem to have been noticed much.

[7] I owe this nice way of putting the point to Ben Jarvis.

on our path towards getting in contact with the external situation, but rather objects of distinct experience. It takes considerable philosophical ingenuity to deny the seemingly obvious fact that feeling something happening in your body is not the same as attending to a dangerous situation. To get around this, Prinz makes statements like 'we feel the offensiveness of external situations resonating through our flesh' (2004: 227). But while I find Prinz's statement attractive, on closer examination the metaphor of resonance suggests that we experience a resemblance between the feeling and the offensiveness of the situation. If so, it entails that we already have independent access to the situational offence, obviating the necessity of bodily feelings. Indeed, as I will argue in Chapter 3, the only way to appropriately trigger those bodily changes would be if the subject does some fairly sophisticated cognitive work first. Thus the key claim of perceptual theory, that bodily changes are needed to reveal the emotional qualities of the situation, is undermined. The cognitive construal seems to be already doing that work for us.

Thus I do not endorse the perceptual theory. However, the perceptual theory made at least one definite advance in emotion theorising. In arguing for the perceptual status of bodily feelings, Prinz opened up the possibility that bodily activity contributes to the intentional function of emotions. This aligns with a more general trend in recent philosophy of mind to explore embodied approaches to cognitive processes. Other contemporary theories of emotion have pursued this potentially fruitful way to synthesise the bodily and meaningful aspects of emotions. For instance, Julien Deonna and Fabrice Teroni (2012) claim that bodily responses are attitudes or modes of representing objects (I discuss this theory further in Sections 3.3 and 4.4). Meanwhile Rebekka Hufendiek (2016) claims that emotions combine information about the situation with information about bodily responses to provide a distinct kind of affordance representation (there are hints of this view in Frijda 1986: 325; 2007: 205). Hufendiek calls this an 'embodied, action-oriented representation' (drawing on a type of representation proposed by Clark 1997). She also raises the interesting suggestion that emotions can be treated as a kind of skill.

The other important feature of Prinz's theory from my point of view is that in order to justify how bodily responses can serve a representational function, he roots the account in a more fundamental theory of mental content. Prinz appeals to Dretske's causal theory of content (to be discussed in the following chapter). As a general strategy, I think that this is absolutely the right way to approach the emotions, and it is surprising that

more philosophers of emotion have not taken this up.[8] It is not enough to simply make comparisons between emotions and other sorts of mental state. We must go back to the basics of mental content and see what resources this provides for connecting up bodily responses and mental content.

## 0.4 The Strategy of This Book

So far, three characteristic features of emotions that we have identified are (1) their intentional nature, (2) the disposition to respond and (3) their evaluative qualities. That is, in emotions, three of the major strands of mind – representation, agency and value – are tightly intertwined. To make sense of this, I do not think we ought to make analogies with other mental states. Instead, I propose that emotions are versions of a distinct sort of mental state that combines all three features in a single basic function. This is what I think valent representation achieves. Thus emotions are not to be treated as the emergent or combined products of other representational states. On the contrary, valent representation is a fundamental type of representational content.

To justify this claim, my goal in Chapter 1 is to motivate the idea of valent representation, independently of any considerations that it may conveniently combine certain features of emotions. To do this, I consider the basic problem of mental content and argue that valent representation is the kind of function that can plausibly emerge out of natural causal processes, while displaying distinctively mental characteristics. I then consider how valent representation may ground other sorts of mental content in which the functions of representation, value and action are not so intimately linked.

Once we establish a new basic sort of mental capacity, we can extend it systematically to cover all manner of mental states, not just emotions. Thus the rest of the book follows the strategy of building on the foundation of valent representation, describing mental states of ever greater sophistication.

My first stop, in Chapter 2, is pleasure and pain. Specifically, I am interested in what lends these experiences their pleasant or unpleasant

[8] A notable exception is Carolyn Price (2015), who grounds her approach to emotions in a teleosemantic theory of mental content (to be discussed in the next chapter). According to this approach 'the content of an emotional evaluation depends on the structure and function of the emotional response that it is supposed to prompt' (2015: 75). Price, however, prefers to embrace the variety of emotional phenomena rather than pin down one essential feature or function (2015: 71).

quality. Following Murat Aydede (2014) I use the label 'affect' to refer to these hedonic aspects of experience. Affect is not limited to pleasure and pain. Hunger, satiety, tiredness, nausea, itchiness, distaste, the pressure to urinate as well as the emotions all generate affect (and thus may be gathered together under the general category of 'affective states'). I argue in this chapter that the simple form of valent representation is insufficient to explain the generation of affect and that we must build an additional higher-order representation on top of valent representation. My model promotes a basically evaluative theory of pain and pleasure, as opposed to imperativist views that some philosophers have recently proposed. However, the higher-order representation I outline also serves a valent function in delivering attentional priority towards the pleasurable or painful object.

In Chapter 3 I finally tackle the emotions. I start by making the case that emotions qualify as a kind of valent representation. Given that valent representation is a general category of mental state, the next order of business is to distinguish the emotions from other affective states like pain or hunger. I argue that the distinctive sophistication of emotions is the way they situate their objects within the wider temporal, modal or social context. Thus my slogan definition is that *emotions are valent representations of situated concerns*. Finally, I explore how the process of association can allow individuals to direct innately given emotion types towards new objects.

In Chapter 4 I turn towards the experience of bodily feelings; the element that somaticists have tended to identify as the essence of emotional states. Because I have already synthesised the representational nature of emotions with dispositions towards bodily activity, I do not need to give the experiential feeling of this activity such a prominent representational role as the somaticists have supposed. I argue for the more traditional view that the intentionality of emotions is distinct from the experience of bodily feelings. However, in recognition of the somaticist position, I argue that bodily feelings do provide a distinct sort of intentional content of their own, which I ultimately analyse as a sense of bodily agency. I argue that this distinct intentional content provides an additional control function for emotional states, allowing certain key differences in emotional types to emerge.

In Chapter 5 I link the representational content of bodily feelings with our representations of the emotions of others, derived via expressive behaviour. I suggest that the same basic capacity to endow a pattern of bodily responses with representational meaning is involved in each case. The important difference is that our expressive interactions with other

people constitute a new level of emotional control – the social emotions. Moreover, I argue that emotional regulation occurs at the social level, rather than at the level of the individual, and is thus a sort of collective cognitive task.[9]

Chapter 6 explores the extent to which we can incorporate our conscious thinking activities under the general purview of the affective control system. I treat our conscious thinking activities as, fundamentally, an extension of the affective response. When we rationally evaluate the situation or make plans, our activity is stimulated by underlying valent representational systems. To this extent, I endorse Hume's (1739/1978) famous claim that reason is the slave of the passions. But at the same time, reasoning is treated as an additional layer of control, serving to both refine and inhibit impulsively generated emotional episodes. Meanwhile, I distance my view from the claim that the *content* of our thoughts or beliefs serves our interests. Yet I do think it's fair to say that we treat our beliefs as a kind of valuable resource. That is, we can be protective about our beliefs, while also willing to 'trade-up' if a more effective or unifying model of the world is on offer. In this way I retain the claim that beliefs aim at truth.

Once reasoning is incorporated into our emotional lives, the complexities of the mature human emotional response are more or less in place. But one final refinement of control is required if we are to get a proper understanding of the ways in which emotions impact our lives.

In Chapter 7 I present an account of character. I begin by distinguishing character from personality. Personality concerns individual variations found at all layers of cognition, while character is more specifically about forming a hierarchy of concerns. My account of character further extends my general control theoretical approach to affective states. I analyse character in terms of long-term sentiments and the ways in which these sentiments make normative demands upon our episodic emotional states – and even our reasoning activities. I argue that forming long-term sentiments relies on a distinctive mental capacity for uniquely tracking objects (particularly people) over time – the narrative capacity.

Chapter 8 reviews the various affective phenomena that I have explored in the book, and in this respect serves as a summary chapter. More

[9] I had originally envisaged tackling collective emotional phenomena at the end of the book, so it was something of a discovery to find it most appropriately explored at this stage in the development of cognitive sophistication. In retrospect, it makes sense that our fundamentally social nature precedes the development of our rational capacities.

importantly, it offers a general control theory of the mind. According to this theory, each distinctive form of affective state (e.g. emotions, feelings, sentiments) corresponds with a layer or level of regulative control. These layers are built on top of each other, allowing the refinement of our regulative activities. Moreover, each new level is characterised by a new representational sophistication, such that descriptive states of mind like perception and belief can emerge as by-products of these affective developments. For this reason I regard the emotions and related affective states as the 'trunk' of the mind, while the different descriptive states are its branches. Alongside this general system of control, I explore our person-level sense of agency, and further ways in which we consciously extend our affective lives by means of environmental interactions, as described in theories of extended cognition.

Finally, in the appendices to this book I offer a glossary of the key terms I employ as well as a summary of an earlier article I published on emotional dimensions in 2009 (with some corrections). Stipulating a set of emotional dimensions is part of the project of differentiating emotions from each other. I show various ways in which the dimensions I outline allow us to make fine distinctions between different emotional states, describe culturally specific emotion concepts and analyse poetic emotional language.

Overall, I hope that this book presents an illuminating new perspective on affective states and their role within the mind as a whole. I consider myself to be building upon and synthesising a large body of insights that has been accumulated by philosophers, psychologists and neurologists over the last few decades. Naturally, there is no general theory of emotions or affective states that I can endorse 100 per cent, since I would not otherwise have written this book. But when I read the work of others in this area, I find myself agreeing more than disagreeing. Similarly, I hope that other philosophers and cognitive scientists interested in these issues will find much to agree with here.