

Precis of *The Emotional Mind*

Tom Cochrane

0. Introduction¹

In this precis I'm going to summarise some of the key ideas in *The Emotional Mind: A control theory of affective states* (Cambridge University Press, 2018). The sections of this precis follow the chapter structure of the book.

The book grew out an attempt to give an original account of the emotions. As anyone who has researched this area can tell you, the more you explore the emotions, the more you realise how complex they are. They connect to all sorts of issues about bodily sensations, actions, values, reasoning, character and social interactions. Ultimately, I don't think a proper account of the emotions is possible without understanding how they are placed within the mind as a whole. This is what I try to provide in the book. As a result, I end up with a theory of mental architecture, where the mind is structured around fundamental life-sustaining processes that deploy ever more elaborate forms of representation to serve their aims.

First some brief background: The history of the emotions debate since the time of William James has been dominated by a debate between the traditional cognitivist position- that emotions are a species of evaluative judgement, and the non-cognitive or somaticist position- that emotions are patterns of bodily sensations. As far as I can tell, everyone these days is searching for the ideal middle ground where evaluative judgement and bodily responses are perfectly synthesized. It is not enough to say that emotions combine bodily responses and judgement. The key adaptive innovation of emotions will lie in how exactly these two features are connected.

In recent times the perceptual theory of emotions has presented a particularly strong synthesis. As Prinz (2004) articulates it, the feelings of bodily changes themselves represent the intentional contents of emotions. However, there have been a number of attacks on this

¹ This precis draws together and expands on material I provided to The Brains Blog (between 11th - 15th Feb 2019) <http://philosophyofbrains.com/2019/02/11/a-new-account-of-the-emotions.aspx> and the Cambridge University Press blog (11th Feb 2019) <http://www.cambridgeblog.org/2019/02/what-is-the-relationship-between-reason-and-emotion/> These sources, in addition to this precis, are citable, though preferably direct citation should be made to the book.

theory, and in my opinion, the perceptual theory is dead. Michael Brady (2013) has a book length attack on the epistemic analogy with perception, pointing out that emotions do not automatically justify belief in the way that perceptions do. It is also clear that emotions are routinely mediated by other perceptual or mental states (like imagining, memory, and the interoception of bodily responses) in a way that just doesn't fit the profile of perception. See Salmela (2011), Deonna & Teroni (2012: Ch. 6), Barlassina & Newen (2013) for further attacks, plus a few in my own book.

A new kind of emotion theory has drawn more deeply on the connection between emotions and action.² It tries to move beyond the perceptual theory by arguing that bodily responses should be understood less as the representational content of an emotional state, and more like its mode or attitude (see e.g. Deonna and Teroni 2012, also Hufendiek 2016 to some extent). That is, we don't represent danger (or loss, or gain); we represent in a threatened, or protective or welcoming mode, by responding in the ways characteristic of fear, sadness or happiness respectively. I do not go quite as far as Deonna and Teroni, since I still believe that emotions represent a particular kind of content. However I similarly regard the evaluative aspect of emotions not as something that we represent, but as captured in an evaluative manner of representing. That is, I claim that it is part of the *representational activity* of emotional states to either increase or decrease the presence of their objects.

1. Valent Representation

This approach to the emotions needs to be motivated. We need a stronger reason to believe in new sort of mental state other than it conveniently combining the intentional, bodily, and evaluative aspects of emotion. The way I provide this motivation is by digging into the problem of mental content (cf. the approach of Prinz 2004).

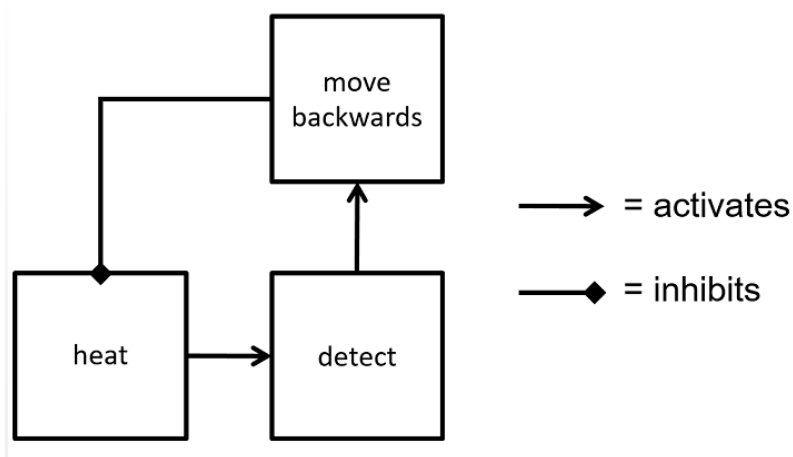
The basic problem of mental content is how some activity inside the head could be about a specific kind of object in the world. Structural resemblances may well be involved, but resemblance is too cheap to fix objects precisely. Thus a lot of philosophers have thought that some kind of pragmatic interaction with the object could fix things precisely. Teleosemantic views are one approach (e.g. Dretske, Millikan). Success semantics is another.

² Nico Frijda is well-known for emphasising this connection (e.g. 1986). However Frijda's account lacks the kind of conceptual clarity that we seek in philosophy.

Success semantics is roughly the claim that a mental state represents whatever object makes actions based on that state successful. So for example, your mental state is about the chair if, on the basis of your mental state, you are able to successfully interact with the chair. In this sense, interaction with the world is necessary for fixing the reference of the mental state.

However a problem for success semantics is that it seems that you need both beliefs/percepts (about the world) and desires (to stimulate actions) to be in place, before *any* mental state can have *any* content at all. That is, to guide an action like sitting down, I need to both perceive the chair and desire to sit down in it. To require both percepts and desires to appear simultaneously is implausibly complex as an account to how mental content naturalistically emerges. To get round this problem, we need mental states that serve the duty of percepts and desires simultaneously.

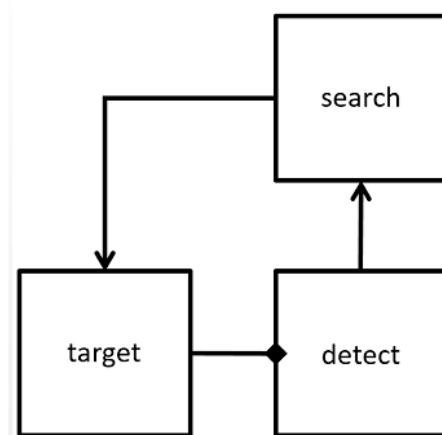
To achieve this, I propose a new fundamental sort of mental state which I call ‘valent representation’. Valent representations make use of negative feedback loops, as depicted below.



An avoidant loop

In this loop, as the level of ‘heat’ gets higher, the activity of the detection node will get higher. This then increases the activity of moving backwards. But if the creature moves backwards, the level of heat (for the organism) will get lower, and so activity in the detection node will fall, eventually leading to inactivity (until the level of heat rises again). Thus the presence of heat for the organism is regulated. I call this an ‘avoidant loop’.

The great thing about control loops like this is that we don't need separate descriptive and directive representations. One representational node (the detection stage in the diagram above) is doing the work of both functions. We can also have 'attractant loops' where if the level of a good thing like food falls low enough, activity in the detector (or hunger) node increases, triggering a response that will secure food, ultimately leading to the inhibition of the feeding response (see the below diagram).



An attractant loop

Now because negative feedback loops are common in nature, there's an issue about how we distinguish the representational ones from non-representational ones, and I spend a lot of time in Chapter One of my book justifying various details, including what makes a creature alive. However, I will skip here to the final definition of valent representation.

X valently represents Y iff:

1. X is a living creature
2. Y is a type of object or property (the object).
3. An indicator³ of Y either activates or inhibits activity in a system internal to X (the detector).
4. The activity of the detector triggers a physically distinct and specific response (the response).
5. The response is capable of decreasing the present capacity of Y to activate the response via the detector (negative feedback regulation).

³ Note, not Y itself. An indicator is a sign of the object, e.g. the colour or smell of some food.

Essentially, valent representation is representation in a valent (i.e. positive or negative) manner. What makes it valent is that the representational activity automatically triggers a bodily response aimed at increasing or decreasing the presence of the object. The theory is still representationalist in the sense that there is an inner state or activity that stands in for the object (i.e. the detection node). But at the same time it incorporates embodied views on cognition, since the response is playing an indispensable role in the *activity* of representation. The object of the mental state (e.g. 'presence of food') is all and only those conditions that the response can regulate.

Finally, an important feature of valent representation is that we can build all sorts of elaborations upon the basic loop structure. For instance, we can make the detection nodes trigger a response only once a certain threshold of activity is reached, or detection can trigger multiple responses, or responses can trigger further responses. Valent representations can also interact with each other, inhibiting and supporting each other. This is how I make sense of the representation of more complex objects. Affective states like pain, pleasure and emotion require more elaborate structures. But ultimately I argue that they all come down to the same fundamental structure of negative feedback control. Indeed, in my more expansive moods I claim that *all* cognition is ultimately an elaboration upon valent representation.

Overall, valent representation elegantly combines the three aspects of the mind: representation, agency, and evaluation. It is a major claim of the book that these three strands start out as merged in the primitive mind, and it is only through various cognitive sophistications that they get teased apart.

2. Pain and pleasure

Upon the foundation stone of valent representation, the rest of the mind is built. My first stop is pain and pleasure.

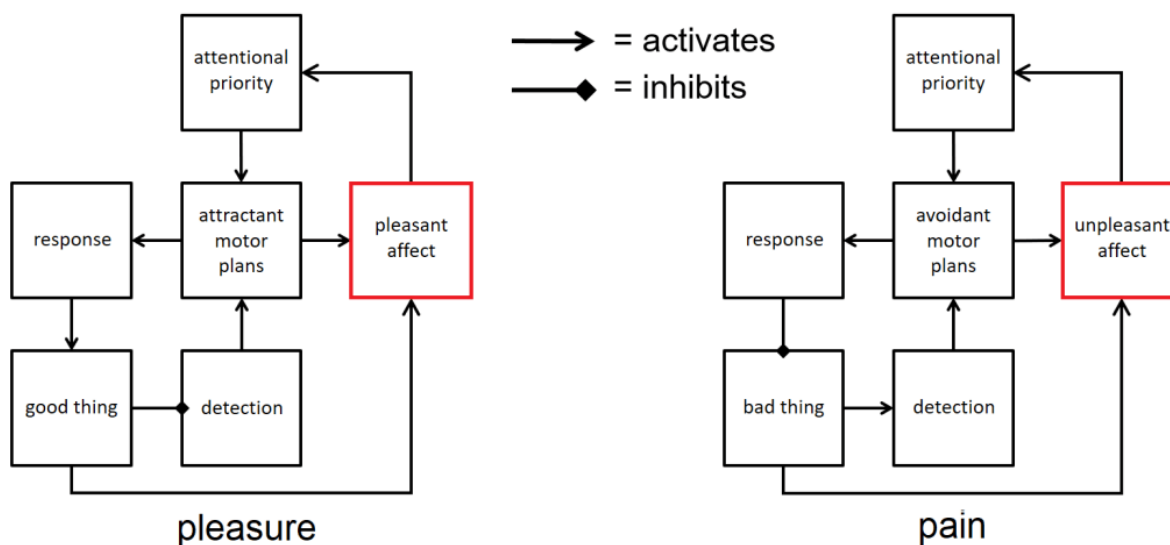
The first thing to note that is although valent representation is representation in a valent (i.e. positive or negative) manner, it is not a representation *of* valence. Indeed, one of the reasons that pain has become such a hot topic lately is due to the recognition that the intrinsically unpleasant aspects of pain are distinct from the representation of damage to the body. A significant justification for this dissociation is the existence of pain asymbolia, which is a

condition in which individuals are able to identify a sensation as a pain, and yet not experience this feeling as bad. The very possibility of this condition forces the distinction upon us. Thus philosophers have been trying to give accounts of what exactly makes an experience intrinsically unpleasant, where some also propose symmetrical accounts for the intrinsically pleasurable aspects of experience. Following Murat Aydede (2014) I use the term ‘affect’ to cover both pleasant and unpleasant aspects of experience. We can then define an affective state as any mental state that involves the arousal of affect. I encourage others to adopt this terminology.

An initial point to make about affect is that there must be some connection with valent treatment. Painful and pleasurable affect must involve approach and avoidance responses respectively (even if we’re just talking about cognitive kinds of response like making plans or recalling details). If they did not, we would have no way to distinguish affect from non-affective sensations, and no way to oppose pain with pleasure. This is not to deny the existence of cases where we might embrace a pain or avoid a pleasure, but those should be understood as complex and derivative cases. Thus valent treatment is necessary for affect.

However, this allows for two options: one where affect is involved in *directing* the valent treatment, and another where affect is involved in *tracking* the valent treatment. These options correspond to two broad approaches to affect: evaluativism- where affect describes the condition of the body (e.g. David Bain) and imperativism- where affect directs the subject to do something (see e.g. Colin Klein).

My approach is basically evaluativist. There are various reasons for this, but one of the major ones is that imperativism does not do well with pleasant affect. This is because a mental state in which one strongly or urgently directs the subject to deliver a certain good is compatible with lacking the good (or the experience of the good) entirely, and such states of yearning are typically quite unpleasant. What we also need for pleasant affect is for such a motivation to co-exist with actually having some of the good thing. Thus, pleasurable affect is essentially a representation of success; that the organism is directed towards acquiring some good, *and* has managed to do so. I call this a ‘success theory’ of pleasure. I then offer a symmetrical ‘failure theory’ of pain, where the subject is directed towards avoiding some harm, and yet the harm is still present. The causal schemata for these states is illustrated below.



Symmetrical models for pleasurable and painful affect

The upshot is that affect is a higher-level representation that combines a representation of harm or benefit, with a representation of the directive response. Thus the model is that on top of regular valent representations there's a higher level node that monitors the presence of the good and the response.

Meanwhile, affect itself follows the structure of a negative feedback loop because it also automatically triggers a response of increased attention.⁴ Again, there are various reasons for thinking that affect directly stimulates attention rather than specific responses, but the main one is that organisms already need to automatically respond to harms and benefits prior to developing the capacities for pain or pleasure.

Anyway, since affect automatically triggers attention, it is an intrinsically motivational state, but because increased attention is common to both pain and pleasure, this response cannot distinguish the two. Rather increased attention indirectly serves a valent function by boosting or prioritising the underlying aversive and appetitive responses.

A final significant aspect of the account is the following hypothesized equation for affect intensity: **Affect = response x object**

⁴ A claim tentatively mooted in the book is that affect is the *only* driver of attention.

That is, the intensity of pleasurable or painful affect is equal to the (detected) intensity of the response multiplied by the (detected) degree to which the harm or benefit is present. This means that if either the object or the response level is detected as 0, the subject should experience no affect. I think this equation could be a significant psychological principle, particularly given the role of affect in directing attention, so I am hopeful that it will be empirically confirmed.

3. Back to Emotions

We can now return to the emotional states. As mentioned earlier, like Deonna and Teroni, I am incorporating the evaluative aspects of emotions into the manner or mode of approaching the object, rather than its representational content. Where my theory differs from that of Deonna & Teroni is that I still attribute a constitutive role to cognitive representation in helping us to grasp the object of the emotion and to trigger the appropriate responses.⁵ That is, if we zoom in on the ‘detection’ box in the negative feedback loop diagrams, we need to make this box a bit more complex to make sense of what’s going on in emotions.

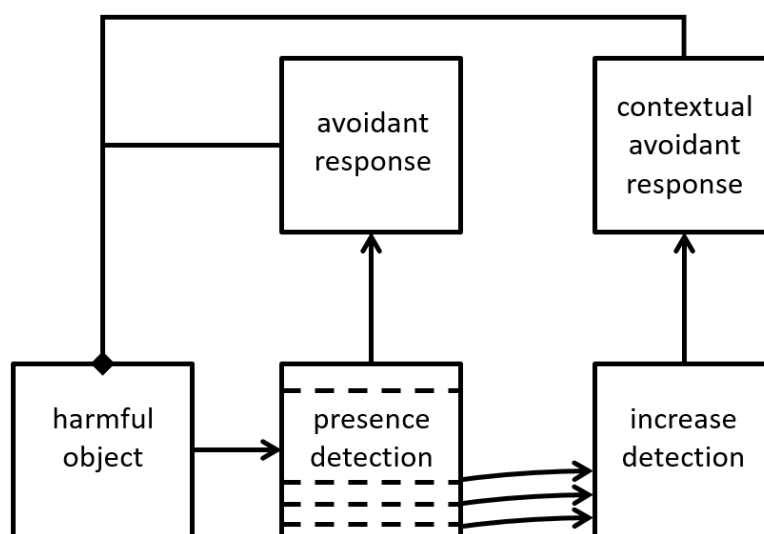
An innovation of mine is to describe the characteristic form that these representations take. I claim that emotions valently represent a contrast between the current state of affairs and some other state not in the here and now. For example, fear involves a representation of an upcoming harm (a temporal contrast), jealousy involves a representation someone else getting a good thing (a social contrast), and regret involves a representation of a bad thing that might not have happened (a modal contrast). Even surprise involves a contrast between expectations and reality (a basic kind of modal contrast). I call these situated concerns, to distinguish them from way that pains and pleasures track immediate harms and benefits. It is plausible that many non-human animals can represent at least some of these contrasts (i.e. by means of memory or imagination) and thus that we can allow for emotions in non-human animals.

It is important to note that the representation of contrasts makes emotions more sophisticated than pains and pleasures because the subject is now oriented towards the wider context. This means that the subject can now serve his or her interests in a more contextually sensitive way, thus making sense of the appearance of the emotional capacity from a selective point of view.

⁵ I share this view with traditional appraisal theories in psychology. In this respect my theory is more cognitive in flavour compared to some recent theories, yet bodily responses are still playing a constitutive role in fixing what the emotions are about, and the evaluative manner the subject regards these objects.

Meanwhile, to represent contrasts only demands cognitive capacities like memory and imagination; capacities that are most likely available to many different species, as well as human infants.

Overall, my slogan definition is that emotions are ‘valent representations of situated concerns’. This is the minimal requirement for all emotional states.



A schema for fear

Above I provide a schematic illustration for the emotion of fear. So in this case there is an underlying aversive negative feedback loop (on the left hand side). On top of this loop is built an additional node that detects when the presence of the harmful object is increasing. When this secondary node is activated, it triggers an avoidant response (e.g. to avoid an area where harms may be likely). In particular, this contextually sensitive response may be triggered *prior* to the avoidant response that would be triggered by the direct presence of the harm. This is because the ‘increase’ node can be triggered by a relatively low level of presence, whereas the direct avoidant response would require a stronger degree of presence. For example, suppose that the creature is sensitive to various signs associated with the presence of a harm, such as a certain smell. This smell is not yet strong enough to stimulate direct avoidance, but the mere fact of detecting an increasing presence is sufficient to motivate pre-emptive avoidance.

Altogether we can (minimally) characterise each emotion in terms of three factors: i) whether the underlying valent representation is positive or negative; ii) the nature of the contrast representation involved (temporal, modal, or social) and; iii) the specific kind of response triggered (e.g. running away versus retching)- which helps to specify the type of object.

4. Coping Potential

A crucial difference between my theory and somaticist theories is that while I think bodily responses play a necessary role in emotions, I do not think the *feelings* of those changes are necessary for emotions. In particular, it is possible to have an emotion where one is unconscious of one's bodily changes. However, I think bodily feelings play a supplementary role in capturing a distinct, though complementary sort of intentional content which I call emotional bodily feeling (see also my 2017 paper, which is an earlier version of chapter 4 of my book).⁶

While emotional bodily feelings are not necessary for emotions in general, they are necessary for some emotions, because they add an additional layer of control. The sophistication here is that by tracking one's automatically triggered bodily responses, and anticipating how well this bodily response can manage one's situation, one can then stimulate further responses that modulate or inhibit one's initial reaction. This is analogous to what psychologists like Klaus Scherer (2005) call 'coping potential'.

In order to grasp the important role that this extra control level has, let us analyse the emotion of sadness. Sadness requires at least three control layers.

First layer: we have an underlying concern. This can be either a positively valent attraction to something (e.g. your beloved), or a negatively valent aversion to something (e.g. being hit). Basic attractions and aversions do not require contrast representations. When the 'level' of the attractant is low, a response is simply triggered to increase the level.

⁶ Indeed my account of emotional bodily feelings resembles Hufendiek's account of emotions, because we both take bodily feelings to represent the capacity of the body to deal with the situation. However, I believe that somaticist theories are overly focusing on this aspect of ordinary emotional experience and failing to properly recognize the work that quite sophisticated cognitive representation is doing in capturing intentional contents, prior to the experience of bodily responses.

Second layer: We represent the attractant as lost (this is a temporal contrast) or an avoidant as gained. This representation automatically triggers a response to reduce the presence of the loss (so it's a negatively valent representation in that sense) by restoring the presence of the lost thing.

Third layer: So far, this response is common to both anger and sadness, so we need a third layer to differentiate anger from sadness. The third layer is where the initial bodily response is represented, but then appraised as inadequate to restore the loss. This leads to a secondary response to inhibit the first set of responses, and to engage compensatory responses instead. That is, sadness is overall the valent representation of being unable to restore a loss.

Meanwhile, anger is where you *do* feel powerful enough to meet the challenge, and you gear yourself precisely to do this- to get back the lost thing, or to remove barriers to getting it. However anger can burn itself out if, by trying, you find you can do absolutely nothing about it- in which case I would expect the response to switch to sadness.

I would also allow the following qualification to the above causal model for sadness. It seems possible that instead of fully triggering the initial restoration response, we simply prepare this response at a neural level. This may be sufficient to generate a bodily feeling, in line with what Antonio Damasio (2000) calls the 'as-if loop' of bodily feeling. It should also be sufficient to enable the individual's coping potential to be calculated, and the response accordingly modulated, prior to triggering useless restoration responses. However, I still claim that it is typical of sadness that an initial restorative response is triggered. For example, in the film *Brokeback Mountain*, there's a very moving scene at the end where Ennis embraces the shirt of his dead lover Jack. Ennis wants Jack, so he grasps at this token, but he's also keenly aware that he cannot get Jack back, and that is the core of his desperate sadness.

A secondary, but I think significant consequence of this theory of coping potential is that it allows for an additional form of value on top of the experience of pleasure. Basically, when one anticipates or judges that one's responses are sufficient to deal with potential demands, one feels confident or powerful. Achieving this feeling can be an important source of motivation.

5. Social Emotions

So far we have four control levels: valent representation, affect, contrast representations and coping potential. To make sense of the social emotions we need even more. Often when we think of social emotions we are thinking of particular emotions like jealousy, contempt, or gratitude. However, we should regard *social emotionality* as an additional layer of regulation that sits on top of regular emotional states. On this view, most human emotions have a social aspect.

To explain this, I once again I appeal to control theory. I claim that social emotions are a valent representation aimed at regulating a relationship. In particular, social emotions involve the regulation of what I call a group's 'affiliative status.' The affiliative status between two or more people is defined as a combination of their power relationship and their intimacy relationship. Our intimacy relationship is fixed by the range of things we are disposed to cooperate on (e.g. from bodily pleasures to paying the bills). Our power relationship characterises what we can expect from each other in terms of support (I might serve you more than you serve me). Thus I may have a relatively intimate and equal relationship with my wife, while I have a relatively distant and power-unbalanced relationship with my boss.

A group's affiliative status can change over time, but crucially, members form expectations about how other members are supposed to act in conformity with it. This includes both the way we treat each other, and the attitudes we might expect ourselves to share towards events in the shared environment (e.g. finding the same things amusing, disgusting or annoying). Thus if my wife behaves in a way contrary to this expectation, I will respond emotionally in a way aimed at restoring the expected relationship. My wife may equally think that my behaviour fails to cohere with our affiliative status, and thus she may respond emotionally in kind. In this way a social-emotional negotiation will take place, and hopefully we will eventually end up on the same page- emotionally speaking. If not, this will probably lead to a change in our affiliative relationship.

The principle mechanism by which social emotions function is via expressions in the face, voice and posture. I largely follow Mitchell Green (2007) in thinking that expressions signal to each other our emotional states for the sake of social negotiation (cf. behavioural ecology views). I supplement this with my analysis of emotional bodily feelings, that is, the

experience we have of the bodily changes triggered in emotional states. Both when I trigger the various bodily changes associated with anger, and when I observe someone else doing the same thing, the same mechanism is involved in taking that pattern of responses to represent the capability and disposition of the agent to act in a certain way. For anger in particular, a reciprocal pattern of expressions allow agents to compare their relative preparedness to attack, and negotiate and establish their affiliative status without resorting to physical violence.

A crucial part of my account is that an affiliative status does not belong to the individual. No individual can unilaterally establish a settled relationship with anyone else. It has to be agreed by all parties involved. As a result, I argue that the regulative process must be understood as a collective phenomenon. I resist calling this social level of regulation a collective emotion (contra Margaret Gilbert 2000). I think we can do everything we need with the notion of social norms. These norms only exist at the social level, and thus there is something irreducibly collective going on. But we need not suppose the existence of group minds or group mental states. These social norms can moreover spread through a population, allowing for the development of cultural-emotional patterns.

In retrospect I should have made more explicit connections in my book to the reactive attitude tradition in ethics, i.e. the idea we get from Peter Strawson that moral relationships are often a matter of feeling and expressing certain emotional states. I talk about social emotionality in larger groups in the book but I only briefly mention the moral community. Note in particular that when the group is larger than a dyad, you can have an emotion aimed at regulating the affiliative status of the group where one group member has harmed another group member, though you are not affected. In this way, my model of social emotions can apply to the resentment or sympathy we might feel for persons we have little contact with, but which are still members of the moral community.

6. Reason and emotion

Reason and emotion are often supposed to be at odds with each other. From one perspective, our emotions are like unruly toddlers, demanding and whimsical, that need to be held in check by the adult intellect. From another perspective, the rational mind is cold and calculating and needs the warmth of the passions to grasp what really matters. I don't think that either of these perspectives, properly understood, is wrong. Where they are potentially

confusing is if they suggest that emotions and reason are two separate sources of agency vying for supremacy. For instance, Plato encourages this confusion with his tripartite model of the soul (e.g. in *The Republic*). Kant does the same in the *Critique of Practical Judgement* (e.g. Bk.I, Ch.3) when he supposes that the rational grasp of moral imperatives can motivate action independently and even in rejection of our passionate impulses.

In contrast to Plato and Kant, we must remember that humans (and other animals) are single agents and we have not evolved the resources of emotion and reason to fight against each other, but to ever more effectively protect the things we care about. Given this consideration, I think the correct thing to say is that reason elaborates emotion. That is, it's yet another layer of control.

Most philosophers and psychologists of emotion agree that emotions have a descriptive function (alongside their motivational function). Emotions inform us about the state of the world- that it is dangerous, enviable, disgusting and so on. Naturally, the kinds of properties we are describing rely on the person caring about certain things; they are relative to the individual in this sense. But given that the person cares about certain things (e.g. the integrity of his body, the status of his loved ones) it can be entirely factual that a situation threatens or supports him. Now along comes the capacity for rational inference. This allows the emotions to massively expand their capacity to track the things the individual cares about, to check whether the initial emotional representation is accurate, to infer consequences, and have further emotions towards those consequences. This, I contend, is the main purpose of reason.

At the same time, it is misleading to say that reason is slave to passions, as Hume famously declares in *A Treatise of Human Nature*. I claim that the motivational juice driving all cognition is drawn from our underlying homeostatic regulation systems. Emotions are one cognitive resource for elaborating these systems while rational inferences are a further resource. This means that one concern-regulating system can overrule another, where the first is rationally elaborated and the second is not.

The resulting experience can be one where we rationally infer that acting impulsively (say to run away from giving an important speech) could destroy one's reputation and we accordingly stop ourselves. Thus there is room for motivational conflict, but nothing so

simple as reason controlling emotion. We could as easily say that one emotion is controlling another emotion here, or that one kind of thinking is controlling another kind of thinking.

7. Character

One final layer of control is required before we have a reasonably complete picture of our emotional lives. This is character. I analyse character as essentially the combination of the individual's sentiments. Sentiments are basically long-term attitudes that we bear towards specific individuals, institutions or causes; that is, the things that we love or hate. The important difference that such sentiments make is in making an individual prioritise the regulation of certain concerns over others, in ways that can allow other emotional processes to be overruled. For example, when I love my children, I prioritise their flourishing over most other concerns. For instance, I may act to protect them even when it compromises my own safety. So the fear for my safety can be overruled by the concern for my children.

The most distinctive cognitive feature of sentiments is that they require us to track unique individuals, not just general types of aversive or attractive situations. Tracking unique individuals is, I argue, our most sophisticated representational activity. In particular, it necessitates that we generate a narrative history about the individual. These narrative histories pick out individuals so uniquely that they could not be replaced by another individual with the same set of qualities. The replacements lack the same history.

Once we understand the way that love uniquely tracks individuals we can make sense of long-term emotions like grief. Peter Goldie uses the example of grief to defend his narrative model of emotions, because he thinks there is no essential emotional response or representation (instead we have different stages of grief that we link together). However, I argue that we can treat grief as a control process in fundamentally the same manner as other emotions. The key complication of grief over sadness is that the underlying concern attaches to a unique individual. Because the attachment is to a unique individual, this makes it impossible for the usual response to either restore or compensate the loss. Thus a slow process is required where the control system has to gradually learn to no longer seek the presence of the loved one. The changes in strategy will correspond to the different stages of grief.

Finally, note that character is not the same as personality. I provide a characterisation of personality in the book in terms of general strategies that individuals develop to regulate their concerns. These strategies standardly combine the sensitivity to a certain concern, say one's social relationships, with capacities, say one's capacity to empathise with others. Like character, personality serves to individuate people, since some people are more sensitive and more capable of regulating certain values than others. Moreover, personality traits are typically stable over the long term. However personality operates at all levels of regulation, so it is not a distinct layer of control in the same way as character.

8. A control theory of the mind

Overall, in my book *The Emotional Mind*, I try to show how the incredibly complex range of activities going on in our emotional lives all fit together. The picture I provide is of one elaboration building on another. This, I think, helps us to see that there is some order in the chaos of our emotional lives.

The final chapter of my book is on mental architecture. This is where I propose a control theory of the mind as a whole. It is perhaps the most ambitious and speculative chapter of a book that is probably already too ambitious for its own good. I am trying to fit together all the various mental processes that I have analysed up to that point (homeostatic processes, pains and pleasures, emotions, bodily feelings, social emotions, emotion driven thinking, and character traits).

At the foundations of the mental architecture I propose lie the valent representations; basic negative feedback control mechanisms that track simple properties and automatically trigger regulative responses. As we evolve more sophisticated ways to represent information, new types of affective state emerge. The basic idea is that new kinds of state elaborate the fundamental control structure, each new level building incrementally upon the last. Thus representational sophistications develop for the sake of guiding our regulative processes more effectively. The more purely descriptive cognitive and perceptual functions are by-products of these developments.

In more detail, I outline the specific levels in the development of mental architecture, corresponding to the different affective states that I analyse over the course of the book. Here

is where the theory is most tentative, but it follows a general logic of how complexity develops step by step. That is, we need cognition of a certain complexity to be in place before cognition of a slightly greater complexity can plausibly emerge.

Key to the development of cognitive complexity, in my view, is the interaction and competition between multiple valent representations. First they inhibit each other. That is, the triggering of one valent representation prevents another valent representation from launching a conflicting response. Then valent representations start to associatively support each other, probably because some properties or the responses they trigger are more mutually compatible than others. Mutually supporting systems are motivated by the need to overcome the inhibitions of competing systems.

The next step in complexity is the formation of higher-level representations. Higher-level representations infer structures or objects behind first-order appearances (cf. connectionist approaches). They go beyond the mere bundles of properties that association captures. I claim that the model of affect I have outlined, in which ‘failure’ and ‘success’ of regulation is represented, is the first higher-order representation. It is higher-order because it represents objects and the individual’s own responses as bearing a relationship with each other. This is not a simple appearance. As mentioned above, affect also plays a special role in assigning attention, thus again serving the competition between valent representational systems.

Emotions then involve a further sophistication, since they involve not just a representation of states of affairs, but a comparison between the current state of affairs and some other state (e.g. in the past or future). Emotional bodily feelings and our awareness of the emotions of others via their expressive behaviour are the next step, involving a simulation of how an agent may behave in the future. This allows new kinds of calculations to be performed where the individual contrasts their emotional state with the emotions of others, or represents them as coordinated in some respect- as in social emotional regulation.

The next steps involve the development of symbolic reference and inference rules. This is of course a major development of mind and even in the book I can only roughly sketch how it fits into the architecture. A basic idea is that once we have social-emotional norms, we are coordinating our references to the world, and we thereby have the beginnings of shared symbolic references to the world. For example, the shared expression of disgust comes to

symbolically reference the disgusting object. However, when it comes to mental architecture, I think a whole new series of developments in representational sophistication build upon the basic foundation of symbolic thinking. I entertain the possibility that the developments in complexity in our symbolic/inferential capacities parallel the prior development of complexity in affective states.

Finally, as described in the previous section, the capacity to generate narrative histories allows us to track unique individuals and to prioritise their well-being over more general or short-term situations.

The full table of steps is provided below. In each case I specify i) the kind of property or object that is targeted; ii) the representational sophistication required to target this property or object; iii) the kind of affective state this object is tracked by; iv) and descriptive states of mind that emerge as by-products (such as regular perception that doesn't directly trigger a regulative response). This precis is too compressed to fully make sense of all these levels, but hopefully it can give readers a sense of how the major functions of the mind might be incorporated in this architecture.

Target	Representational capacity	Affective State	Descriptive by-products
Single qualities	Valent representation (VR)	Valent representation	--
Conflicting qualities	Inhibition	Inhibited VRs	Perception
Clustering qualities	Association	Clustered VRs	Mental imagery and memory
Regulative success	Higher-order representation	Affect	Object-types, patterns
Temporal / modal context of object	Change / contrast representation	Emotions	Object persistence through change
Response patterns	Action simulations	Social emotions / Bodily feelings	Representations of minds
Convergent responses	Representation of norms	Collective emotional norms	Symbolic reference
Inferred objects	Inference / norms of thought	Emotions with cognitive responses	Inference rules
Unique objects	Narrative histories	Sentiments	Singular thought

Control architecture of the mind

The book ends with an appendix on emotion dimensions, which concerns the key ways in which we distinguish emotions from each other. Seven dimensions are outlined: valence, power, temporal orientation, probability, social connectedness, temporal duration, and generality.⁷ The reader may note that all of these concepts have been mentioned at one point or another in the above precis. The combination of these dimensions allows us to differentiate extremely finely between emotion types and even to map emotional nuances in art and literature that lack linguistic labels in English.

References

Aydede, M. (2014). How to Unify Theories of Sensory Pleasure: An Adverbialist Proposal. *Review of Philosophy and Psychology*, 5(1), 119-133.

Bain, D. (2013). What makes pains unpleasant? *Philosophical Studies*, 166(1), 69-89.

Barlassina, L. & Newen, A. (2013). The Role of Bodily Perception in Emotion: In Defense of an Impure Somatic Theory. *Philosophy and Phenomenological Research*, 89(3), 637-678.

Brady, M. S. (2013). *Emotional Insight: The Epistemic Role of Emotional Experience*. Oxford: Oxford University Press.

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT press.

Cochrane, T. (2017). The double intentionality of emotional experience. *European Journal of Philosophy*, 25(4), 1454-1475.

Cochrane, T. (2009). Eight Dimensions for the Emotions. *Social Science Information*, 48(3), 379-420.

Damasio, A. R. (2000). *The Feeling of What Happens: Body, emotion and the making of consciousness*. London: Vintage.

⁷ This summarises an earlier publication of mine (Cochrane 2009), though in that article I outlined 8 dimensions, and now there is only 7 (I dropped the dimension of freedom).

Deonna, J. A. & Teroni, F. (2012). *The Emotions: A philosophical introduction*. Oxford: Routledge.

Frijda, N. H. (1986). *The Emotions*. Cambridge: Cambridge University Press.

Gilbert, M. (2000). *Sociality and Responsibility: New essays in Plural Subject Theory*. Maryland: Rowman & Littlefield publishers.

Goldie, P. (2012). *The Mess Inside: Narrative, emotion, and the mind*. Oxford: Oxford University Press. Green, M. S. (2007). *Self-Expression*. Oxford: Oxford University Press.

Hufendiek, R. (2016). *Embodied Emotions: A Naturalist Approach to a Normative Phenomenon*. New York: Routledge.

Millikan, R. G. (1995). Pushmi-pullyu representations. *Philosophical perspectives*, 9, 185-200.

Prinz, J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press.

Salmela, M. (2011). Can Emotion Be Modelled on Perception? *Dialectica*, 65(1), 1-29.

Scherer, K. (2005). What are emotions and how can they be measured? *Social Science Information*, 44(4), 695-729.

Strawson, P. F. (1962). Freedom and Resentment. *Proceedings of the British Academy*, 48, 1–25.