Emotions as Transitions

1. Introduction

One task of psychology is to identify cognitive capacities. We have a capacity to parse sentences, to identify faces, to sort things by their colour, and so on. Capacities serve as theoretical data – they are observed phenomena that one then seeks to better understand. One thing we might wish to understand is how these capacities are implemented. Take sentence parsing. Is this something achieved through some compositional process or, more like Natural Language Processing AI (*e.g.* GPT-3), by predicting the next word based on the previous words and training data? What structures underly *our* ability to parse sentences? There are likely to be many candidates, at least in principle.

In order to uncover the inner workings of our capacities, we look to 'effects'. Most of us have the capacity to distinguish between spoken 'ba' and 'fa' sounds. One thought is that this is achieved through aural sensitivities that detect changes in vibration picked up by the eardrum. But the McGurk Effect suggests that there is more to the story. Without changing the incoming vibrations, sound experience can be modulated by showing a video of a mouth making a 'ba' sound or a 'fa' sound with a consistent sound overlaid.[1] We learn that our overall auditory experiences are at least in part determined by visual cues in addition to what's first picked up by our eardrums. The McGurk Effect gives us a hint into the inner workings of audition and helps us better understand the capacity to discriminate sounds of a certain sort.

In the present paper, the focus is on emotional capacities and a well known effect – recalcitrance. Recalcitrant emotions, such as fearing the dog even though one knows that the dog is harmless or being angry with one's partner even when one realises it was only in a dream that the partner was nasty, have played the role of effect in much theorising about emotions. But in my view, we've stayed a bit too close to home, aiming

[1] See McGurk and MacDonald (1976).

to fit the effect into a paradigm – the representationalist paradigm – that isn't fit for purpose. I will use this criticism as a launching off point to introduce a different way of thinking about emotions that is better suited to making sense of recalcitrance. I will argue that emotions are transitions between representational states rather than being representational states themselves. The view is better suited to make sense of recalcitrance and, at the end of the paper, I will offer reasons for thinking that main points that speak in favour of a representationalist approach to emotion can be recaptured or explained away by the transitions view.

2. Recalcitrant Emotions

Recalcitrant emotions conflict with our considered judgements. Faced with the rollercoaster deemed entirely safe, one might still feel afraid. After a vivid dream, one might feel angry with their partner and yet judge that, in reality, their partner did nothing wrong. Recalcitrance is an effect. We have the capacity to engage with situations emotionally and this is a capacity that can be suppressed or even absent. We want to better understand this capacity – how do we *do* it? The phenomenon of recalcitrance shows us something about the inner workings of our emotional systems. Focus for a moment on fear. Creatures that are sensitive to threats might have been structured in such a way as to easily and reliably shut down threat reactions once it is recognised that there is in fact no threat. But that's not how it goes with us, at least not all of the time. One who is afraid of spiders can't typically read that the spider approaching them is of an entirely harmless variety and then happily let it crawl onto their hand. If one is serious about getting over the fear, one must not only learn about spiders but must handle them in a controlled environment, learn to 'sit' with one's fear as they are in the presence of a spider, and so on.[2] Our emotions have a kind of lasting effect and a recalcitrance to change. At least some of the time. Sometimes emotions are relatively easy to 'turn off': *I wake up in a panic – classes started five weeks ago so today is the big exam! I completely forgot to study. I may as well skip a morning coffee because I'm wide awake now! But wait, it's a long weekend! There is no exam today. It's still a*

[2] See McLean et al (2022) for a meta-analysis of work on exposure therapy.

week away. It's the sixth meeting *of the term, not the sixth week. Thank goodness!*. In but a moment we go from stricken with emotion to calm again. This too needs explaining. Emotions are recalcitrant… sort of… sometimes. When and why hasn't been satisfactorily teased out. But all of this is useful data.

Recalcitrance, in the philosophical literature, is a complicated matter. There are at least two aspects of recalcitrance that have garnered attention and they aren't always properly separated. First, there is the *recalcitrant* aspect of recalcitrant emotions. Some emotions don't want to budge. We might know they aren't called for, but that knowledge doesn't seem to make a difference. It's not that the emotions *cannot* be changed, but they show a stubbornness to change. And it is has been noted that this stubbornness looks to be different from change in run-of-the-mill beliefs. We can quite easily change from believing it is sunny outside to not believing it just by asking a friend who just came in from the park. In light of this, we might think that in some ways emotion looks more like perception than belief – when gripped by a visual illusion, there is no talking yourself out of it. Let's call this aspect of recalcitrant emotions the 'stubbornness effect'. The stubbornness effect, at least on first glance, looks to speak against a belief theory of emotion and more in favour of something like a perception theory (but more on this shortly).[3]

Second, there is a rational or normative aspect to recalcitrant emotions.[4] When one judges that the glass walking bridge is stable and safe but trembles and recoils as one steps onto the clear surface, the emotion/judgement pair creates a *normative* conflict – one *shouldn't* both be afraid and judge that there is nothing dangerous, one *shouldn't* be angry with her partner while judging that she did nothing wrong. Such pairs are structurally inconsistent in the sense that they exert a kind of rational pressure on one

[3] See Majeed (2019), Tappolet (2012, 2016), and Prinz (2004). As noted, the issues aren't always sharply separated, though these authors highlight the apparent encapsulation of emotion.

[4] This aspect of recalcitrance has been discussed in considerably more detail amongst philosophers. For recent discussion that has refined the normative issue see Benbaji (2013), Brady (2007, 2009), D'Arms and Jacobson (2003), Döring (2008, 2009, 2015), Greenspan (1981, 1988), Grzankowski (2017, 2020), Helm (2015), and Majeed (2020).

another that is relieved when one of the attitudes is revised. Let's call this aspect of recalcitrant emotions the 'normative data'. I resist calling it an 'effect' since it isn't the sort of thing I'd expect psychologists or cognitive scientists to try to isolate in specialised settings. But even so, it is a theoretical datum. In the present paper, my focus is on the stubbornness effect. In section 5 I'll have a little to say about the normativity data since it is important that the view I offer not conflict with it.

3. Representational theories

Since the 1960s, a dominant philosophical conception of the emotions has it that they are evaluations of value.[5] To be afraid of the dog, for example, is to evaluate the dog as dangerous. To be angry with the loud neighbour is to evaluate the person as having wronged you. The stubbornness effect has played an important role in shaping and refining this approach to the emotions.

Once it is recognised that emotions have intentional objects – my anger is *directed upon* you, my fear *at* the dog, and my envy *at* the suave person with the Aperol at 2pm – and that those objects are taken to be some way – the dog is taken to be dangerous – a very natural thought is that the emotions are *value judgements*. According to Cognitivists or Judgementalists, at least part of what it is to be in an emotional state is to be in a state of belief or to make a judgement. Judgements or beliefs are not mere causes of emotions, they are constituents.[6]

But recalcitrant emotions pressure this picture. Beliefs and judgements do not display the stubbornness that we would expect if the Cognitivist theory were correct. I believe that today is Tuesday. You show me on your phone that it is in fact Monday. Now I believe it is Monday. I judge that there is a red cup on the table. You show me the red film that has been placed over the lights and assure me the cup is white. I then judge

[5] See Kenny (1963). For a discussion of the emergence of this this kind of approach see Deigh (1994).

[6] See Solomon (1976), Neu (2000), and Nussbaum (2001).

that, despite appearances, the cup is in fact white. This is not to say that we never have stubborn beliefs or are never stubborn in making judgements, but belief and judgement show a kind of malleability in the face of incoming information that emotions often do not show. So when we aim to better understand our emotional capacities by hypothesising that *what they are* are beliefs or judgements, we run into the stubbornness effect which speaks against this hypothesis.

But nothing in the foregoing speaks against the broader idea that our emotions are evaluations. That is, that they are representations as of things instantiating value properties. Perceptions are often deemed representational but they aren't malleable in the same way as beliefs or judgements. The famous example involves appeal to the Müller-Lyer lines. The lines appear to be different in length, but when we measure them, we come to know they are the same in length. But this makes no difference to how things look and the illusion persists. With this visual effect in hand, Perceptualists can offer a refined version of the evaluation theory of emotions. On this view, a constituent of our emotions is a perception or appearance of value.[7]

But Perceptualism doesn't quite align with the stubbornness effect either. I've been shown the Müller-Lyer lines a hundred times at least and never has my knowledge that they are the same in length made one bit of difference to how they look. But I had a dream the other day that my father was uncharacteristically mean and belittling. I felt angry until about noon and then the anger wore off. A friend signed up for an intensive CBT course to get over a fear a spiders and it was surprisingly effective – they can now scoop one up with a cup and paper and release it outside whereas a month prior they would sooner sell the house than deal with a spider. Exposure therapy doesn't make visual illusions go away. And sometimes, as in the case earlier of the big exam, emotions can be *quickly* undercut. None of this quite aligns with perception. I suspect you can provide plenty of anecdotes of your own. So the stubbornness effect is again a problem.

[7] See Döring (2008), Prinz (2004), and Tappolet (2012).

There might appear to be a relatively easy way forward. Hypothesis: *some emotions have a cognitive component and some a perceptual component*. But one worry about this sort of disjunctive approach is that it posits a joint where we might expect unity. I can felicitously say, 'I am afraid of the big exam and the spider'. Isn't my ability to elide 'afraid' just ahead of 'the spider' some evidence that I'm talking about the same kind of thing in being afraid of the exam and the spider? But perhaps this linguistic data shouldn't carry too much weight. The world is a complicated place and what might look to be one thing sometimes turns out otherwise. There is a more pressing concern facing the disjucntive view.

On the disjunctive view we should expect that some emotions are relatively easy to change – as easy as coming to believing that it is sunny outside – and some are not things we can change with new information – as immovable as a visual illusion. But there is a third category, many emotions are somewhere in the middle. Return to the exposure therapy and the spider. Although not always effective, CBT utilises exposure therapy with reasonable success in many circumstances and what one finds is a gradual change over time. Similarly for anger management. One doesn't learn a new fact about an individual who wronged them and then happily go on their way. Rather, over extended time, one learns strategies that reduce anger reactions and works towards a more general change in how one approaches situations. This takes practice and time and is a bit more like learning a skill than it is like learning a new fact or like getting a pair of glasses. The data on the ground is that emotions are sometimes easy to change but not always, and yet they aren't completely immovable but tend not to be moved with mere information. This all needs accounting for and can be used to guide us.

4. Emotions as Transitions

Let's put aside for a moment the cases where emotions seem immovable and those in which they are relatively easy to change with incoming information – the Perceptualists and Cognitivists have, respectively, something to say about those cases. I'll come back to

those cases later. At the moment, let us focus on the cases where change is elicited through practice and exposure. These cases fall between the cracks of the going theories.

First pass, we might simply add yet another disjunct to our theory. New hypothesis: *some information states are open to incoming information, some are encapsulated, and some are encased in a permeable membrane.* I confess I don't have a knockdown argument against this kind of approach, but I think it is striking that the information states – belief, memory, and perception – that have garnered a lot of attention amongst psychologists and cognitive scientists don't seem to have this semi-permeable feature. Of course we might simply posit it given the observed effects, but I want to offer a different kind of suggestion.[8] Namely, that emotions aren't further representational states. They aren't in the family of memory stores, beliefs, or perceptions. Rather, emotions are the kinds of things that can be shaped and moulded over time and with practise because they are something else: transitions between representational states which are themselves underwritten by mental architecture. In slogan form, emotions are part of the hardware rather than the software.

An analogy will help bring out the idea. Suppose you have a simple computing machine and a tape with the value TF on it and you hope to get back a tape with T on it. Your truth functional machine gives back Ts and Fs depending on the Ts and Fs it gets as inputs. When you put in TF you get back back an F. You discover that one way to get back a T instead is by giving it TT rather than TF, FF, or FT. Your little machine seems to be following the AND rule and so you know you can get back T only if you send in TT. But there is another way to get a T rather than an F as output. You could keep your original TF as input, crack open the machine, and rewire it. What you find inside is an AND chip, a chip that gives back a T iff it receives TT. But with some physical tinkering, you could replace or modify the AND chip with an OR chip and then when you feed in TF, you will get back T as desired. Writing different values on the input tape is relatively

[8] Mitchell (2021) argues that emotions are *sui generis* states, not to be modelled on perceptions or cognitions. Gendler (2008) invents a new category – 'aliefs' – but I worry this labels the issue rather than illuminates it.

easy in my imagined scenario (all you need is an eraser and pencil) but rewriting the machine is a good bit harder. But both will get you the desired result of an output of T.

I don't want to take the hardware/software or the simple computer analogy overly literally. I don't wish to subscribe to the view that our minds are *just like* simple computers. But I do think the analogy is helpful in various ways. The representational theory of emotion that is the philosophical mainstay really is committed to representational states – specifically states that represent things as instantiating value properties. But any representational theory is going to need, in addition to representations, some mechanisms that *do something* with representations. Suppose we take perception to be an information gatherer that feeds into belief. It's no good simply having incoming information and a storage container. Somehow we need to get what's coming in (perceptual information) into the storage container (memory and belief). And along the way we should expect some changes – resolving visual ambiguities, bringing things under concepts, and so on. Similarly for reasoning. We believe that p and believe that if p then q, and we draw the conclusion that q. In addition to the beliefs, we make a *transition* from premises to conclusion.[9] So in order to draw inferences we need some kind of transition system (whether you think that's like a simple computer or something else entirely). Fodor (1975) held that there could be no computation without representation, but equally important is the idea that representations are pretty useless without some way of manipulating them and using them. My suggestion is that emotions can and should be accounted for on the manipulation and transition side of the equation rather than on the representation side.

This view allows us to better capture the recalcitrance effect but before turning to that, the idea needs to be spelled out with more care.

[9] Cummins (1982) and Pylyshyn (1991) for general discussion. See Quilty-Dunn and Mandelbaum (2017) for an application of transition to inference. See Johnston (2020) for an application of this idea to bias.

Emotions are things that happen. They are events. We experience bouts of fear and anxiety and become angry with friends *for a time*. My suggestion is that emotion events are events of transition from incoming representational states to further representational states and bodily changes. To take a simple example, suppose you round a corner on a trail and you spot a coiled snake. The pattern on the snake is one with which you are familiar and you take the snake to be venomous. You feel fear come over you, you want to turn back, and you can feel your heart racing. On the view I'm suggesting, your perceptual and cognitive inputs are manipulated resulting in a desire to flee, the judgement that the snake is a threat, and a racing heart. That process of manipulation *is* your episode of fear and it is a process that might continue for a duration.

Psychologists should find the outlines of this idea familiar. The Affect Program Theory of emotions is very much in keeping with the idea I'm suggesting.[10] Unfortunately, in my view, advocates of the Affect Program Theory tie their view to the idea that the affect programmes are innate mechanisms and much of the criticism of these views is targeted at claims concerning innateness and universality. The view I'm offering needn't take any stand on this status of the mechanisms and their transitional activities – they might be innate or might not be, might be the product of social construction, or might be determined by genetics. Structurally speaking, these are decisions that are not forced upon us by holding that emotions are transitions.

How do we individuate these transitions? A forthcoming way is to look for the typical triggers and downstream effects. On the view on offer, there are underlying dispositions which are triggered in particular ways and which trigger particular things. The project of

[10] See Ekman and Cordaro (2011), Izard and Malatesta (1987) and Tomkins (1970). Moors (2022, p. 116) provides a useful example that brings out the structure of the Affect Program Theory: 'To illustrate this sequence, the processing of a crouching tiger activates the affect program for fear, which consists of a link between the representation of the tiger and the tendency to flee. This, in turn, generates an adrenaline rush and blood flow to the legs (mobilizing the body in general and preparing for flight), a fear expression (in face, voice, and/or gestures), actual flight behavior, and the feeling we call fear'. I take issue with the very last thought. On my view, it is the activated affect program that is experienced (more on this in section 5).

individuating emotions then becomes the project of providing those dispositional profiles. For example, fear is triggered by perceptions of dangerous things and beliefs about threats and leads to desires to fight or flee, judgements to the effect that things are dangerous, and bodily responses such as an arched back and increased heart rate (no doubt that's overly simplistic, but hopefully one sees the approach). The events of undergoing emotional episodes are not themselves dispositions but are rather the processes of those dispositions being engaged. But if you want to know *what fear is*, for example, the way in is to pin down the typical causes and effects that trigger the disposition. An episode of emotion is the process of transitioning typically kicked off by the identified cluster of causes and which typically leads to the cluster of identified effects.

One might worry that the causes and effects will be too coarse for this individuative task as we might find that many of the same causes and effects are present in, say, both fear and excitement. The representationalist alternative is that there is, *in the content of the emotion*, a difference in evaluative property and on these grounds the representationalist might claim an advantage. But the view I'm offering can take advantage of this very feature. The transitions view of emotions takes it that the inputs to the transitions are cognitive and perceptual states, and one is free to individuate those triggers just as finely as the representationalist does. The relevant difference is that the emotional episode itself doesn't represent, but, of importance, we will look to the representational inputs and outputs in addition to behaviour. So the worry that the typical causes and effects don't individuate finely enough can't be successfully leveraged by the representationalist. (More in a moment on the cognitive bases of emotion.)

There will be some further worries to ward off later, but before turning to those, one should wonder what speaks in favour of this revisionary position. I want to offer a preliminary consideration in favour of this view and then return to recalcitrance. I don't want to rest too much on the preliminary point, but it is worth airing.

The transition view avoids a kind of representational redundancy. Everyone (Neo-Jamesieans, Cognitivsts, Perceptualists) in debates over what emotions are seems to be prepared to agree that emotions have 'cognitive bases'. Either as constituents or as causes, emotions are 'based on' perceptions and/or judgements. It is, at least in part, because one perceives things to be a certain way or judges them to be a certain way that one undergoes an emotional episode and the typical representational story has it that emotions get their contents from their cognitive bases. But let suppose now that one judges that the snake before them is dangerous (the same points can be made if one takes the cognitive base to be a perception). One now has a representation to the effect that the snake is dangerous in their mental economy. The representationalist takes on the burden of explaining why it is that we need to, as it were, represent that the snake is dangerous *again*. That's already represented, so why represent it again? I admit there are possible replies here, but I do think the representationalist takes on an explanatory burden the transition theory avoids. As a first reply to this redundancy worry, one might pressure the claim of 'again', noting that on sophisticated representational views the emotions have the cognitive bases *as constituents*. But this doesn't quite get around the worry for there remains a question about why exactly we should think that a state that has a representation as a constituent represents what its constituent represents – that inheritance inference doesn't come for free. We can then ask again, 'why bother with such an inheritance view when you already had the representation in the base?'. As a second reply, one might argue that although the content of the emotion is the same as or similar to the content of the cognitive base, the content features in the mental economy in a different way (much in the way one might suppose that p and then come to believe that p). I think this is a much more promising way of avoiding the charge of redundancy but I will leave it to my opponent to fill out the details. At some point we need to do something with our representations. Why not take the emotions to be the doing rather than more representing?

As noted, I don't want to rest too much on the redundancy worry. It's the recalcitrance effect that really speak in favour of the transitions view.

Returning to recalcitrance, we saw three kinds of cases we wanted to make sense of.

A. Emotions that behave like belief in being very open to change by information.
B. Emotions that behave like perceptions in being very stubborn to change by information.
C. Emotions that change, but through processes such as training and exposure rather than information change.

The transitions view of emotion looks to both the cognitive bases and the transition to make sense of A-C. Let 'F' be the disposition that is triggered when one undergoes a bout of fear and let 'eF' designate the event of transitioning from inputs to outputs via that disposition being triggered. Candidate inputs are perceptions ('P') and cognitive states ('C'). For present purposes we needn't focus on 'O', the outputs (though they are a crucial part of the individuative story).

When one is undergoing an emotional episode, we have a structure that looks like this:

P/C —> eF —> O

Now how might we make changes to eF? One thing we might do is undercut the triggers.

Suppose we are looking at the case of fear and that the the trigger is a perception. You can stop a perceptual trigger (seeing a snake, say) by putting a wall between you and the snake, by closing your eyes, by turning around, and so on. In reality this might not undercut all emotional triggers since you might have formed a memory or a judgement about the snake and closing your eyes won't make those go away, but let's keep the example simple for the moment. You put a wall between you and the snake and you thereby undercut the trigger and begin to calm down. In another scenario, you take five steps back from the glass bridge and focus on the stable, opaque land beneath your feet and feel the fear recede. As the trigger is a perception, being *told* you are in no danger won't undercut this perception, and so new information coming in may not help disrupt eF. The perception needs to be cut off and being told the snake is harmless isn't a good way to accomplish that. My explanation of B, then, is that perceptions are amongst the

triggers of emotions and perceptions aren't changed in the relevant ways via incoming information such as testimony.

But suppose the trigger is a belief or a judgement. These *are* sensitive to incoming information. I might believe that there is venomous snake just outside the door and become afraid on this basis. I fully trust you and you tell me there is no snake. As a result, I change what I believe and undercut this trigger of eF. This type of approach accounts for A.

So far, these are cases that the representationalists can cover as well. The hard case for that view was a case like C. The transition view adds a layer of structure that allows us to say more than the representational views. Perhaps the input is not undercut – I'm still faced with the snake or I still believe it to be a threat. Nevertheless, I might alter the disposition that these inputs tend to trigger much in the way that I suggested above that I might swap the AND chip for an OR chip. As this is not a process of information change – neither the disposition nor the event of it being triggered and unfolding is a representational state – we have some purchase on why it is malleable but in a different way from perception or belief. The dispositional changes are architectural changes that take more time and come about in different ways than information changes. This is difficult for representationalists to capture in anything but an *ad hoc* way – in terms of a representational state that is somehow and somewhere in the middle between belief and perception, somewhat flexible and somewhat permeable. It's not that this is an obvious non-starter, but what the recalcitrant effect suggests to me is that there is a different kind of mechanism at play in cases that fit the C profile. Information change isn't the right model for C. The transition view better explains the effect (and it has an elegant way of capturing A and B to boot).

5. Representationalism's Revenge?

5.1 Transitions just are *representations*

A representationalist might agree with the transitions view and then reply by arguing that the transitions view really *just is* a representational view. To see how this might go, imagine that one holds that a chip that takes TT to T but all else to F *represents* AND. That is, that the chip either has a representational content expressed by 'if input TT, output T, otherwise output F' or the chip is taken to 'tacitly represent' AND. But I think these ideas are a mistake.

First, Lewis Carroll (1895) has powerfully argued that this way of thinking leads to regress. What Carroll famously shows (or at least a well received lesson from Carroll's discussion)[11] is that we must differentiate writing a rule down from following a rule. If we have a system that computes the AND function, what we need is a mechanism to get us from, say, TF to F or TT to T. We don't need a further *premise* and adding one merely adds more information that eventually we must do something with. So the point from earlier about having both representations and ways of doing something with them kicks in again.

Second, if we insist that our system that takes TT to T and all else to F *tacitly* represents AND, we run into two problems. The first is that I worry we reach a mere verbal dispute. If the representationalist agreed with the shape and structure of the transition view but labels a transition a 'representation*', the only quibble I have is a misleading label. If representing and tacitly representing (representing*) aren't the same thing, we seem to agree on all but the terms.

The second problem with the tacit representation idea is that positing a representation is a substantive matter and we should not trivialise appeals to representation in explanation. This is an important point, so I'll expand on it.

What exactly is required for some entity to be a representation is a well worn and controversial issue, but for our purposes a comparison between two cases is enough to

[11] See Besson (2018) for further discussion.

make the desired points.[12] Consider the gear in a watch that controls the minute hand. As the gear turns, the minute hand turns and the gear was *designed* to co-vary in position with the minute hand. It's possible for the gear to come loose and for it to spin freely without the minute hand moving and so to be faulty. And yet, despite meeting many of the classical markers of being a representation (co-varying, being designed to do so, and possibly malfunctioning), it is intuitively implausible that the location of the gear *represents* the location of the minute hand. No doubt, it's easy to read one fact off of the other when things are going well in the watch, but being able to be easily transform something into a representation isn't the same as being a representation. A bit of mercury sitting in a puddle on a rock doesn't represent temperature even though it easily could be recruited to do so. So being a co-varying thing isn't enough to be a representation and even being a co-varying thing that was designed to co-vary isn't yet a representation. But compare the gear and minute hand case with another:

> Foraging desert ants, Cataglyphis fortis, return to their nest by keeping a running total of their distance and direction from the nest. This mode of navigation was called path integration by Mittelstaedt and Mittelstaedt (Mittelstaedt and Mittelstaedt, 1982), who provided the first systematic studies of this phenomenon, and vector navigation by Wehner (Wehner, 1982; Wehner, 1983). More recent reviews and considerations on path integration are given elsewhere (Wehner and Srinivasan, 2003; Merkle et al., 2006). By path integration the ants acquire a home vector that enables them to return at any time along the beeline, so to speak, to the nest. However, after having played out their home vector, they do not always arrive exactly at the entrance of their nest, a tiny hole in the desert floor. […] If an ant fails to find the nest after having 'run off' its home vector, it terminates its almost straight inbound run and starts a systematic search for the nest (Wehner and Srinivasan, 1981). During this search C. fortis performs loops of increasing radius around the supposed nest position (Wehner and Wehner, 1986). (Merkle *et al* 2006 p.3545)

[12] See Ramsey (2007), especially chapter 5.

When all goes well, the ant finds food and heads straight home. But if, by mother nature or by experimental design, the ant is moved off track, the ant arrives at what should be home, but home isn't found. The ant then initiates a search.

Focus on the ant's relation to the nest. When the nest isn't found, something must explain not only why the ant begins its systematic looping search (and it can't be the nest itself since the nest isn't where the ant is!) but also why that looping ceases when the nest is found. A very attractive explanation is that the ant has something like a placeholder, 'home', and when, using its running total of distance and direction, it fails to arrive at home, that placeholder is utilised. There is some way of checking 'am I home?', there is a time when that checking is to occur, and then the is the initiation of a search when the answer is 'no!'. Finally, the ant ceases the search only when the answer is 'yes, I'm home' (and does not for example cease when finding a white pebble or a leaf).

As others have argued, in addition to being designed to co-vary with other things, representations serve as stand-ins in situations of absence[13] and they serve to guide further processing and behaviour. This is exactly what we find with the ant and what differentiates the case from the gears of the watch. Representations carry information that is utilised by a system and that is available when the worldly correlates are missing.

On the view I am advocating, emotions are not bearers of information but are rather users and manipulators of information. Now of course, none of this shows that I'm *right* about this. Rather, the aim of the foregoing is to show that if one comes along with me in thinking that emotions are transitions, there are good reasons to then resist that they

[13] See Newen and Vosgerau (2020) and Orlandi (2020) for further discussion.

are also representations.[14] If one wishes to call them tacit representations, so be it, but what the AND-chip is doing in our earlier example looks importantly different from serving as a stand-in in situations of absence.

5.2 If emotions aren't representations, they aren't about anything

A key advantage of the representationalist view is that allows us to explain the apparent intentionality of the emotions. Intuitively, my anger is *directed at* the politician, your fear is *of* the snake, my envy *concerns* your success, and so on. Does the transitions view have to deny these claims?

On the view I am advocating, judgements about the aboutness of the emotions are to be accounted for in terms of the triggers and effects of the emotional transitions. If I am angry with Henry, it is Henry representations (be they perceptions, judgements, or something else) that trigger the transition and the transition leads to desires concerning Henry, actions related to Henry, and so on. This is not to say that any old Henry-representation is a necessary and sufficient condition for kicking off the transition and nor is it a way of aiming to develop something like a conceptual role semantics. My point presently is that we are licensed in our aboutness talk by the fact that the triggers and effects of the transition that *is* my fearing of Henry have a common subject, namely Henry. As I suggested above, the emotions will be individuated in a fine-grained way by appealing to quite specific representational triggers and effects.

[14] Given the way I'm thinking about representations, whether emotions are representations becomes a testable hypothesis. If one were to design a clever, controlled experiment, one should, as cognitive science is in the business of doing, be able to isolate and manipulate mechanisms that support or pressure the hypothesis that there is information in a system that is being used in downstream processing and/or to guide action and behaviour. For example, if the representationalist posits that in fearing the snake one represents the snake as dangerous, one should (at least in principle) be able to find evidence that speaks in favour of there being snake-information and dangerousness-information in the system beyond what's present in belief or perception. I'm under no delusions: this will be very difficult to test for. But positing representations becomes a testable hypothesis when one thinks of them as suggested in the main text.

5.3 There is something it is like to emote, but dispositions and the triggering of them aren't like anything

It might now be worried that the view cannot account for the phenomenology of emotion. When one undergoes an emotional episode, there is something it is like. I agree and I hold that there is something it is like to transition from information states to further information states and actions. For one who is wedded to a representational view of conscious experience, this will not be satisfying. But I see no reason to be hostage to that commitment. There is, intuitively, something it is like to partake in actions and activities: it is thrilling to jump from a high cliff, it is painful to run flat out for a long duration, it pleasant to rub a sore neck. Some events are phenomenal episodes and I see no reason to deny that the process of transitioning from incoming representations to further representations and actions has a phenomenology. There is, on my view, something it is like for mental activity to occur – there is something it is like to transition from inputs to outputs. That seems neither more nor less mysterious than the claim that there is something it is like to represent something. Neither theorist has a solution to the hard problem of consciousness, but what could it be about a process of transition that *disqualifies it* from being conscious?

5.4 Emotions justify and can be justified and this shows they have content

How, on the transitions view, could emotions be the sorts of things that are justified by beliefs and that can justify downstream thoughts? One might think that only something with content such as a perception or a belief could play such roles.

To this I offer three replies. First, one might be skeptical of the data at the outset. It is not at all obvious that emotions are justified or that they justify. I don't want to rest too much on this since I do think we talk as though emotions stand in epistemic relations and I think we can accommodate this fact, but there certainly is a long tradition of thinking that emotion is the enemy of rationality and not a participant in it.

My second reply is that we should question whether justifiers must be representational. Plenty of philosophers are non-representationalist about perception, but they do not then go onto deny that perceptions justify beliefs. This doesn't yet help with the idea that emotions themselves can seem to be justified, but it pressures the general idea that justification explanations rely on content explanations.[15]

My third reply is my most serious. I suggest that the epistemic facts in question can be accounted for by the cognitive bases of the emotions. If, as is widely agreed, emotions are based on states such as perceptions, memories, judgements, and beliefs, we can account for our epistemic judgements by appealing to them. For example, suppose that I have no good reason whatsoever to believe that you wronged me but I go on believing it anyway. On that basis I come to be angry with you. To the extent that we want to apply epistemic blame to the emotions, we will want to say that this is unjustified anger. But the cognitive base is also unjustified. So we can account for the epistemic using the cognitive base. And this approach has a further advantage. Suppose you agree that perceptions can justify but they themselves are not open for epistemic criticism. Consider someone feeling afraid while standing on the sky bridge over the Grand Canyon. Unlike the unjustified anger based on the baseless belief that you wronged me, it seems plausible to hold that this is *not* a case where we'd criticise the acrophobic for being irrationally afraid. If the cognitive base is also not open to criticism, we have a story to tell about why it is that some emotions come in for criticism and others don't.

This third reply might be met with the following concern: there are cases where an emotion seems to justify a belief but where there is no good candidate cognitive base to serve as justification; the emotion itself does the work. Take the well-known example of Huck Finn and Jim. Going beyond Twain a bit, let us stipulate that Huck simply has no beliefs about the moral status of slavery. He thinks his travels with Jim are illegal but he simply hasn't thought about the moral issues. But he imagines his friend being arrested

[15] See Raleigh (2017) for a detailed and critical look at the view that whenever a mental state rationally justifies a belief it is in virtue of inferential relations holding between the contents of the two states.

and placed back into slavery and feels disgust at the idea. On the basis of this emotional reaction he judges, for the very first time, that slavery is repugnant. This judgement looks to be both caused by and (crucially) epistemically supported by his emotional reaction. It doesn't look like a cognitive base will do the work, so isn't my suggestion of relying on the cognitive base in trouble in such a case?

This is an important concern. A flat footed reply is that there *is* a cognitive base to do the work: it is his imagining of his friend being mistreated that triggers the disgust. But I think I can offer a more robust reply. Transitions themselves have normative properties. This is a familiar idea when thinking about inference. When one makes the inferential transition from *if p, then q*; *p* to the conclusion *q*, one makes a *rational* transition. The transition itself is a praiseworthy process. In outline, my suggestion is that in the case of Huck, it's not that Huck's judgement is *based on good evidence* but rather that it is the result of a praiseworthy transition. This is not to say that there is a perfect analogy with a rational inference, for in the case of Huck there doesn't seem to be a clear case of moving from a content that logically relates to another content as in in the *modus ponens* example. The critical point is that we are comfortable evaluating transitions normatively. More generally, processes and events where representation isn't obviously present are evaluated normatively. Actions can be *fitting* and *unfitting*. It is fitting to hug a loved one upon arrival but it is unfitting to hug a stranger on a train. But it isn't at all obvious that these actions are themselves representational.[16] So if we take it as a data point that it is fitting to fear dangerous things, but unfitting to fear harmless things, holding that emotions are transitions rather than representations in no way undercuts an application of fittingness to emotion. Nonrepresentational things such as actions are candidates for being fitting, so why not transitions? Transitions are evaluated for rationality, so why not fittingness too? With this idea in hand, I submit that it is fitting for Huck to feel disgust when imagining abuse and it is, in turn, fitting to then believe abuse is wrong. It is fitting to *transition* from imagining abuse to the judgement that abuse is wrong. This is what *Huck* gets right. So much like our discussion of the

[16] See Naar (2022) for more on fitting actions and for a defence of the view that emotions are more like actions than they are like cognitions or perceptions.

stubbornness effect, the transitions view gives us an extra piece to play with. We can appeal to the normative properties of the cognitive bases but also to the properties of the transition itself. It looks like we really need this extra structure given the landscape of cases and the transition view delivers it.

5.5 Emotions normatively conflict with judgements and beliefs and this structural tension must be due to a contradiction in content

In setting up the present paper I noted that recalcitrant emotions display both a stubbornness effect and display normative conflict. When one judges that the glass walking bridge is stable and safe but trembles and recoils as one steps onto the clear surface, the emotion/judgement pair creates a conflict – one *shouldn't* both be afraid and judge that there is nothing dangerous, one *shouldn't* be angry with her partner while judging that she did nothing wrong. One forthcoming way of making sense of this is to hold that the content of the judgement *contradicts* the content of the emotion.

But with fittingness in hand from the previous section, this structural conflict can be accounted for by the transition theorist. If one is angry with Ffion and judges that Ffion has done nothing wrong, then one's anger is appropriate or fitting just in case one's judgment is false making it inappropriate or unfitting. And the reverse is true as well. If one's judgment is true and so fitting, then one's anger will be unfitting because the true judgment that Ffion did nothing wrong entails that the object of one's anger hasn't done anything wrong. The normative correctness of the judgment turns on whether Ffion did something wrong because judgements ought to be true. The normative correctness of the anger turns on the whether Ffion did something wrong because anger is fitting when one has been wronged (and as we just saw, this is compatible with the transitions approach). The structural normative pressure exerted when judgements conflict with emotion is that, necessarily, one violates a norm when one is angry with that which is judged not to have wronged one. Crucially, one needn't appeal to contradictions in

content to capture the conflict.[17] *Conflicts in fittingness* provide an explanation and are consistent with the non-representationalist commitments of the transitions view.

6. Conclusion

By examining the stubbornness effect, I have offered the outlines a different way of thinking about emotions from the philosophical norm: emotions are transitions rather than further representations. The view provides the structure and differences from representational states that are needed to account for the various ways in which emotions can be stubborn and can undergo change. The view also retains a home for cognitive bases and so still has representational resources at its disposal and the view appeals to a mechanism that any good representational theory will want to make use of anyway: mechanisms that manipulate and engage with representations. In the penultimate section of the paper I considered a number of representationalist worries and showed how the proponent of the transitions theory can reply to them. I hope the foregoing provides one with reason enough to explore the transitions view of emotion as a non-representational alternative.

[17] See Grzankowski (2020, 2021) for a detailed discussion of this kind of approach to the normative tension in cases of recalcitrance.

Bibliography

Benbaji, H. (2013). 'How is Recalcitrant Emotion Possible?' *Australasian Journal of Philosophy*: 1-23.

Besson, C. (2018). 'Norms, reasons and reasoning: a guide through Lewis Carroll's regress argument'. In Daniel Star (ed.), The Oxford Handbook of Reasons and Normativity. Oxford, Oxford University Press.

Brady, M. (2007). 'Recalcitrant Emotions and Visual Illusions'. *American Philosophical Quarterly* 44 (3):273 - 284.

—(2009). 'The Irrationality of Recalcitrant Emotions'. Philosophical Studies 145 (3):413 - 430.

Carroll, L. (1895). 'What the tortoise said to Achilles'. *Mind* 4 (14):278-280.

Cummins, R. (1982). 'The Internal Manual Model of Psychological Explanation'. *Cognition and Brain Theory*, 5(3), pp. 257–68.

D'Arms, J. and Jacobson, D. (2003). 'The Significance of Recalcitrant Emotion (or, anti-quasijugmentalism). Royal Institute of Philosophy Supplement, 52, 127–145.

Deigh, J. (1994). 'Cognitivism in the theory of emotions'. *Ethics*, 104(4), 824–854.

Deonna, J. A. and Teroni, F. (2012). *The emotions: A philosophical introduction*. London: Routledge.

Döring, S (2008). 'Conflict Without Contradiction' in Georg Brun (ed.), *Epistemology and Emotions*, Dominique Kuenzle und Ulvi Dogouglu, London: Ashgate 2008, pp. 83-104.

—(2009). 'The Logic of Emotional Experience: Noninferentiality and the Problem of Conflict Without Contradiction'. *Emotion Review* 1/3: 240–47.

—(2015). 'What's Wrong With Recalcitrant Emotions? From Irrationality to Challenge of Agential Identity'. *Dialectica*. 69 (3):381-402.

Ekman, P., and Cordaro, D. (2011). 'What is meant by calling emotions basic?' *Emotion Review*, 3(4), 364–370. https://doi.org/10.1177/1754073911410740

Fodor, J. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.

Gendler, Tamar Szabó (2008). 'Alief and Belief'. *Journal of Philosophy* 105 (10):634-663.

Greenspan, P. (1988). *Emotions and Reasons: an Inquiry into Emotional Justification*, New York: Routledge, Chapman and Hall.

—(1981). 'Emotions as Evaluations'. *Pacific Philosophical Quarterly* 62: 158-69.

Johnson, Gabbrielle M. (2020). The Structure of Bias. Mind 129 (516):1193-1236.

Grzankowski, A. (2017). 'The Real Trouble with Recalcitrant Emotions'. *Erkenntnis* 82 (3):641-651.

— (2020). 'Navigating Recalcitrant Emotions'. *Journal of Philosophy* 117 (9):501-519.

— (2021). 'A puzzle for evaluation theories of desire'. *Thought: A Journal of Philosophy* 10 (2):90-98.

Helm, B. (2015). 'Emotions and Recalcitrance: Reevaluating the Perceptual Model'. *Dialectica* 69.3, pp. 417–33.

Izard, C. E., and Malatesta, C. Z. (1987). 'Perspectives on emotional development: Differential emotions theory of early emotional development'. In J. D. Osofsky (Ed.), *Handbook of infant development* (2nd ed., pp. 494–554). Wiley Interscience.
Kenny, A. (1963). *Action, Emotion, and Will*. London: Routledge.

Majeed, R. (2019). 'What Can Information Encapsulation Tell Us About Emotional Rationality?' In Laura Candiotto (ed.), The Value of Emotions for Knowledge. Palgrave Macmillan. pp. 51-69.

—(2020). 'What Not to Make of Recalcitrant Emotions'. *Erkenntnis*.
McGurk H., MacDonald J. (1976). 'Hearing lips and seeing voices'. *Nature*. **264** (5588): 746–8.

McLean, C. et al (2022). 'Exposure therapy for PTSD: A meta-analysis'. *Clinical Psychology Review*, Volume 91, 102115, ISSN 0272-7358.
Merkle, T., Markus K. and Rüdiger W. (2006), 'Uncertainty about nest position influences systematic search strategies in desert ants', The Journal of Experimental Biology 209, 3545-3549.

Mitchell, J. (2021). *Emotion as Feeling Towards Value: A Theory of Emotional Experience*. Oxford, UK: Oxford University Press.

Moors, A. (2022). *Demystifying emotions: A typology of theories in psychology and philosophy*. Cambridge, UK: Cambridge University Press.

Naar, H. (2022). 'Emotion: More like Action than Perception'. *Erkenntnis* **87**, 2715–2744. https://doi.org/10.1007/s10670-020-00324-2

Neu, J. (2000). *A tear is an intellectual thing: The meaning of emotions*. Oxford: Oxford University Press.

Newen A. and Gottfried Vosgerau, G. (2020). 'Situated Mental Representations: Why We Need Mental Representations and How We Should Understand Them', in Smortchkova, Joulia, Schlicht, Tobias & Dolega, Krzysztof (eds). *What Are Mental Representations?* Oxford University Press. Pp.178-212

Nussbaum, M. (2001). *Upheavals of thought: The intelligence of emotions*. New York: Cambridge University Press.

Orlandi, N. (2020) 'Representing as Coordinating with Absence', in in Smortchkova, Joulia, Schlicht, Tobias & Dolega, Krzysztof (eds). *What Are Mental Representations*? Oxford University Press. Pp. 101-134.

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

Pylyshyn, Z. (1980). 'Computation and Cognition: Issues in the Foundations of Cognitive Science'. Behavioral and Brain Sciences, 3(1), pp. 111–32.

—1991: 'Rules and Representations: Chomsky and Representational Realism'. In Asa Kaher (ed.), The Chomskyan Turn, pp. 231–51. Oxford: Basil Blackwell.

—(2016). Emotions, Value, and Agency. Oxford University Press.
Raleigh, T. (2017). 'Against an Inferentialist Dogma'. *Synthese* 194 (4):1397-1421.

Ramsey, W. (2007). *Representation Reconsidered*. Cambridge University Press.

Solomon, R. (1976). *The passions*. New York: Anchor Press/Doubleday.

Tappolet, C. (2012). 'Emotions, Perceptions, and Emotional Illusions'. In Calabi Clotilde (ed.), *Perceptual Illusions. Philosophical and Psychological Essays*. Palgrave-Macmillan. pp. 207-24.

Tomkins, S. S. (1970). 'Affects as primary motivational system'. In M. B. Arnold (Ed.), *Feelings and emotions: The Loyola symposium* (pp. 101–110). Academic Press. https://doi.org/10.1016/b978–0-12-063550-4.50013-9

Quilty-Dunn, J. and Eric Mandelbaum (2017). 'Inferential Transitions'. *Australasian Journal of Philosophy*, 96(3), pp. 532–47.