Synthese (in press)
https://doi.org/10.1007/s11229-020-02972-x

Pathological Prediction: A Top-down Cause of Organic Disease

Elena Walsh

Received: 2 March 2020 / Accepted: 23 November 2020

> We see what we want to see, what we expect to see, instead of what's really there. I don't think we do it on purpose, most of the time.
>
> Lauren Miller

Abstract Though predictive processing (PP) approaches to the mind were originally applied to exteroceptive perception, i.e., vision and action, recent work has started to explore the role of interoceptive perception, i.e., emotion and affect (Seth, 2013; Wilkinson et al, 2019; Miller and Clark, 2018; Barrett and Simmons, 2015; Van de Cruys, 2017; Barrett, 2017). This article builds on this work by extending PP beyond emotion to the construction of emotional dispositions. I employ principles from dynamical systems theory and PP to provide a model of how dispositional anger (also known as 'hostile attribution bias' or HAB) can develop in response to early experiences of psychosocial stress. The model is then deployed to explain the established link between psychosocial stress in early life and the appearance of certain organic disease phenotypes (such as cardiovascular disease) in later life. This phenomenon can appear mysterious when viewed through the standard biomedical explanatory lens, which has difficulty accounting for the causal influence of subjective perceptions and evaluations of the social and material environment on the development of organic disease processes. The model provided presents such cases as instances of developmental mismatch. They occur when an organism develops an emotional disposition that leads them to make habitually-biased appraisals of what the social environment affords. The model provides a novel explanation of certain organic disease phenotypes with top-down and developmental causes, and demystifies one class of cases involving apparently spooky 'mind-to-matter' causation.

Keywords Dynamical systems theory; Emotion; Emotional development; Life history theory; Predictive processing; Top-down causation

Acknowledgements. Thanks are owed to the members of the Theory and Method in Biosciences Lab at the Charles Perkins Centre, University of Sydney, for discussion of the central

Elena Walsh
The University of Wollongong
E-mail: ewalsh@uow.edu.au

ideas in this manuscript, including Pierrick Bourrat, Axel Constant, Caitrin Donovan, Wesley Fang, Stefan Gawronski, Paul Griffiths, Kate Lynch, Maureen O'Malley, and Arnaud Pocheville. I am also grateful for insightful and constructive feedback received from anonymous reviewers. This feedback was instrumental in developing the final line of argument appearing herein.

1 Introduction

A vision of the mind as a multi-level probabilistic prediction engine has been gaining ground in recent years. According to this picture, what we perceive and experience in the world is not the result of passively processing sensory input. Instead, it is the result of a (constantly) active inferential process that operates in accordance with a Bayesian approach to probability: sensory input constrains estimates of prior probability (from past experience) to create the posterior probabilities that serve as 'beliefs' about the causes of input in the present (Clark, 2016, 2013; Barrett and Simmons, 2015). The brain works to minimise prediction error over time by adjusting predictions to match sensory input, or skilfully engaging in the world such that future sensory input comes to more closely match predictions. According to this picture, what we see and experience is a combination of what we expect to see and what the world actually provides.

Though predictive processing (PP) approaches to the mind have primarily been applied to exteroceptive perception, e.g., vision and action, recent work has started to explore the role of interoceptive perception (Seth, 2013; Wilkinson et al, 2019; Miller and Clark, 2018; Barrett and Simmons, 2015; Van de Cruys, 2017). This work takes as its starting point the James-Lange tradition of identifying emotional states with the perception of changes in the body as a response to stimuli (i.e., interoception) and seeks to integrate interoceptive and exteroceptive predictive processes. Beyond its contribution to 'embodied' approaches to understanding the mind, this work has also generated novel approaches to explaining the genesis of certain psychological disorders, including depression and anxiety (Barrett and Simmons, 2015) and PTSD (Wilkinson et al, 2017).

This article builds on such approaches by providing a conceptual model accounting for the established link between psychosocial stress in early life and the appearance of certain organic disease phenotypes (such as cardiovascular disease) in later life. Over the past thirty years, a growing body of evidence has shown that prolonged psychosocial stress – especially in childhood – can somehow 'get under the skin' in a way that persists across multiple decades and influences risk for disease in midlife. In a watershed study, Felitti et al (1998) showed that 'adverse childhood experiences' (such as domestic violence, abuse, or having a parent in prison) were strongly predictive of disease phenotypes in later life. The study showed, for instance, that people who experienced more than four adverse childhood experiences were twice as likely to be diagnosed with cancer than those who had not faced any form of childhood adversity, and that, for each adverse childhood experience an individual had, their chance of being hospitalised with an autoimmune disease in adulthood rose by 20 per cent.

More recent evidence suggests that the connection between the early psychosocial environment and later disease is strongly tied up with a person's subjective (and perhaps idiosyncratic) interpretation of their social environment. The study of human social genomics has found that a person's perception of the conditions of their social environment (as hostile

or threatening, say) affects the activity of not just a few genes but entire gene profiles that influence susceptibility to disease (Slavich and Cole, 2013). At the level of physiology, it has been shown that the perception of an environment as hostile or threatening leads to changes in hypertensive status and systolic, diastolic, and ambulatory blood pressure (Dolezsar et al, 2014). When sustained over time, these changes increase susceptibility to cardiovascular disease.

When seen through a certain lens, these correlations appear to be doubly mysterious. The first issue is that, in biomedicine, explanations of the causal mechanisms that underpin disease are, when they are provided at all, typically skewed toward a reductive physicalism. Describing disease means describing biological dysfunction, and the presence of dysfunction is causally explained by identifying the biological and physiological mechanisms that underpin it (Carel, 2011; Svenaeus, 2013; Ongaro and Ward, 2017). There is no straightforward way to accommodate a top-down, non-physical cause within such a framework. So we are left with a phenomenon that lacks an explanation – one that might, if we emphasise the role of subjective perception, appear to involve a 'spooky' connection between the mind and the body and troubling claims about mental causation (famously problematic for reasons discussed by Kim (2006, 1999). The second issue is that we are dealing here with 'long-distance' correlations, i.e., ones that link experiences in childhood to disease phenotypes that present decades later. Any mechanism we posit to explain the connection must be one that works over long distances.

This article outlines a novel proposal for explaining the link between early instances of psychosocial stress (e.g., abuse, violence) and the later development of hypertension and cardiovascular disease. It proposes that stressful early life experiences direct the formation of habitual emotional response patterns that persist into adulthood, even once the individual is far removed from the original stressful environmental context. I focus on the case of dispositional anger, characterised, as the name suggests, by a tendency to feel anger or hostility, even when the outer circumstances do not seem to warrant it. There are myriad descriptions of this sort of individual – thin-skinned, touchy, the sort who 'flies into a rage' or among whom one must 'walk on eggshells'. There is also a name for such a disposition in psychology. Influential work by Dodge and Somberg (1987) suggested that repeated experiences of abuse, such as aggression and/or violence lead children to develop a 'hostile attributional bias' (HAB) that filters subsequent experience. Dodge claimed that those who make hostile attributions will, when exposed to a frustrating social stimulus, such as being hit in the back with a ball, tend to interpret the stimulus as an aggressive cue and thus respond aggressively. Dodge and Somberg (1987) called this a sort of habitual 'cue distortion' (also see Anderson and Graham, 2007). Sections 3 and 4 outline a mechanism that explains how early experiences of psychosocial stress (such as abuse or violence) set up the conditions for HAB. Section 5 outlines how HAB disrupts physiological allostasis, eventually producing functional and structural abnormalities consistent with cardiovascular disease presentation. The product is an explanation of how a subset of cases of hypertension and cardiovascular disease are strongly influenced by the development of an emotional disposition skewed toward habitual anger.

Before proceeding further, Section 2 provides some necessary background. After very briefly situating the work to come in the recent literature connecting emotions to PP, it proposes grounding an explanation of how mature emotional dispositions develop in an account of emotions as both dynamical and coordinative: dynamical in the sense of being phenomena that occur across some measurable and transient period of time, and coordinative in the sense that they involve (and are indeed identifiable with respect to) the transient coordination (or synchronisation) of emotion components.

2 Emotion: Coordinative and dynamical

There is a growing consensus that emotion and cognition are inextricably intertwined (Colombetti, 2014; Pessoa, 2013; Miller and Clark, 2018), and recent work connecting emotional processes to PP is at the helm of this development. Existing proposals have extended the idea that exteroception and proprioception involve prediction to interoception. This approach either relates emotion to, or identifies emotion with, interoceptive predictions about sensations from the internal milieu of the body, e.g., heart rate, glucose levels, temperature (Seth, 2013; Barrett and Simmons, 2015; Van de Cruys, 2017; Wilkinson et al, 2019). Now, making predictions about what the body is feeling is not just an internal matter – interoceptive predictions also inform the system of what is happening in the environment. Barrett and Simmons (2015) for instance, propose that the visceromotor cortices generate autonomic, hormonal and immunological predictions to adjust how the internal systems of the body deploy autonomic, metabolic and immunological resources to deal with the sensory world as the brain predicts it to be. Miller and Clark (2018), building on the neuroanatomical evidence of Pessoa (2013) and others, introduce a predictive architecture in which affective information is constantly being fed to brain regions that modulate vision, specifically the medial and lateral pulvinar. They propose that these regions integrate various streams of information including affect, action, value, and attention to amplify emotionally-salient sensory input (a proposal I discuss further in Section 4).

Interoceptive predictions have also recently been used to explain the basis of psychopathologies such as PTSD (Wilkinson et al, 2017) and depression (Badcock et al, 2017; Barrett and Simmons, 2015). These proposals appeal to the role of affect in biasing prediction generation. For example, PTSD can be presented as a system with a learning history of trauma leading to sustained hypervigilance and anxiety that biases perceptual inference to tend to favour threatening hypotheses over benign ones (Wilkinson et al, 2017). Depression can be seen as a system which, due to continual negative feedback concerning rewards eventually comes to predict that rewards are unavailable in the world (Badcock et al, 2017). These accounts rely on connecting a certain life history (e.g., involving trauma, or involving the failure to acquire rewards) to physiological features that linger in the body, that in turn facilitate the development of a habitually-biased form of perceptual inference (connections I will also rely on to explain the development of HAB). In these accounts, the system's interoceptive predictions (e.g., affect) feed into higher-level predictions about what the current environmental stimulus affords.

Emotionality is centrally involved, then, in making higher-level predictions, at least inasmuch as affect is centrally involved. But, actually, an emotion proper is, at least on the dynamical and coordinative proposal I am going to adopt here, different to affect. The latter is always part of the former but it is, as I am about to argue, a coordination or synchronisation of emotion components (including but not limited to affect) that is distinctive of an emotional episode. I emphasise this point because, though the case has been made for emotionality in general being centrally involved in hypothesis generation, something quite distinctive occurs during a window of emotional synchronisation that may not occur during other sorts of affect-mediated hypothesis generation (or, at least, not to the same degree). Before I expand on this point (in Sections 3 and 4) I introduce the two key ideas central to the conception of emotion I use as a basis for explaining the development of emotional dispositions. The first is of emotion playing a fundamentally *coordinative* or organisational function – creating predictable patterns of change in multiple subsystems of the body, mind, and nervous system. The second is of an emotion as an episode of *dynamical* activity, i.e., of continuous change and development over a period of time. It is obvious that emotions are phenomena that occur

over time: the point of calling them dynamical is to emphasise the relevance of the sequence of changes that make up the pattern, in particular the relevance of time-indexed modeling of these changes, when it comes to identifying causal mechanisms that enable the development of emotional dispositions over ontogeny.

So to start with coordination: An emotional episode can be defined as a certain pattern of coordinated changes that occur across a short interval of time among subsystems that serve allostasis in the organism. For example, a fear response could be characterised as a pattern of coordinated changes comprising a threat appraisal, fearful facial affect, elevated sympathetic arousal, and an urge to flee a threatening situation (Hollenstein and Lanteigne, 2014). An anger response could be characterised as a pattern of changes in which different measures for physiological arousal (such as heart rate and palm skin conductance) all become elevated for the duration of the emotional episode (Herrald and Tomaka, 2002; Stemmler et al, 2007). This view can be understood as rather uncontroversial: it is simply one way to cash out the intuitive idea that an emotional episode is a pattern, i.e., a pattern of coordinated changes that occurs across time.

Indeed, the idea that emotion coordinates or synchronises certain subsystems of the organism fits neatly in the story of emotion's adaptive function, i.e., that what is functional in the organisational capacity of emotion is that it galvanizes the organism to expediently produce an adaptive response to an environmental stimulus. The view that emotion is adaptive in this way represents a consensus view in the psychobiological literature on emotion (Ekman and Cordaro, 2011; Ekman, 1999; Griffiths, 1997). At least in its canonical incarnation, this approach is preformationist in spirit. According to Ekman's work in the seventies and eighties (Ekman, 1980, 1973; Ekman et al, 1972), developmental outcomes are 'pre-planned' and executed via a series of static mechanisms. The problem with the preformationist approach is that it has difficulty explaining individual variation in emoting habits. Now, emotion researchers are generally not unsympathetic to the idea that developmental outcomes being seen as 'innate' or 'pre-planned' will result in a failure to explain individual variation (see, for instance, (Morag, 2016). But what is almost never discussed is how to actually explain how emotional dispositions develop over time in response to environmental stimuli, and this brings us to our second idea of emotions as *dynamical*.

A somewhat recent development in the field of emotion theory is the suggestion that we attempt to gain traction on the nature of emotion using dynamical systems theory (see Lewis, 2005; Scherer, 2009a; Colombetti, 2014). There have been, for instance, recent attempts to identify emotional episodes with attractors (for discussion see Colombetti, 2014; Scherer, 2009b) and for a computational model, see Meuleman (2015). An attractor is a point or region on the state space that the system's trajectory frequently visits. The trajectory of the system is determined by the differential equations that govern it, while these equations are themselves based on an initial data-set that charts the changes in each of a set of chosen variables relative to time (i.e., time functions as an independent variable). In the context of emotion the variables typically chosen to build the model are those that comprise a set that displays a covariational pattern thought to be characteristic of particular emotion category types (anger, fear, etc.). Evidence favouring the existence of such covariational patterns (sometimes called 'concordance' or 'synchronisation' in the literature) is robust and has been found for a number of distinct emotion types (see, for instance Hollenstein and Lanteigne, 2014). Time-series data shows that these patterns form over a short period of time, appearing in a matter of seconds (Bulteel et al, 2014), and in many cases as quickly as 600 to 800 milliseconds (Lewis, 2005).

The work of Colombetti (2014) has been especially valuable in providing a broadly enactivist framework within which we can speak of emotions and emotion-related phenomena

as attractors in a dynamical system. However, her approach does not, as far as I understand, speak specifically to the question of how an emotional disposition is generated. In particular, it does not provide a causal mechanism linking instances of a given emotion type to the generation of a longstanding emotional disposition. There have also been recent proposals to use dynamical systems to understand how emotional dispositions develop (Lewis and Liu, 2011; Lewis, 2005). The idea here is that a dynamical approach to emotion can ground a representation of an emotional disposition as a non-temporary attractor region on the state space that represents the dynamical system (a person we 'walk on eggshells' around has a stable and strong attractor for the emotion of anger, a person we think of as habitually timid has a stable and strong attractor for fear, etc.).

Adopting a dynamical approach to modeling emotional dispositions enables us to characterise a disposition as a temporal extension of a covariational pattern that matches a particular emotion category (a disposition toward anger, for example, is a temporal extension of episodes of anger featuring the distinctive covariational pattern associated with this emotion category). It has been proposed (Lewis and Liu, 2011; Lewis, 2005) that such an approach will enable us to unite real-time explanation and developmental explanation, in turn providing a basis for explaining the formation of emotional dispositions. A linchpin of this approach is the claim that emotional episodes heighten Hebbian learning in neural networks. Lewis (2017) suggests, for example, that Hebbian learning can explain how repeated instances of, say, sadness, lead to an increase in synaptic connections that promote sadness on future occasions.

As I see it, PP may have richer explanatory resources than those present if we limit ourselves exclusively to Hebbian learning.[1] In particular, I think PP is better able to address the central problematic under consideration here, namely cases where an emotional disposition yields a habitual tendency to misinterpret signals in the social environment. In particular, PP has the capacity to explain this feature without denying a conception of emotions as functionally Janus-faced: as alerting us at once to the presence of demeaning offenses, while also biasing our attentional and cognitive capacities to privilege perceptual inputs and thought processes that support the hypothesis that the environment affords a demeaning offense over those that might disconfirm it. This dual functionality means that anger can alert us to the presence of genuine hostility in the social environment, but can also yield false positives. In Section 4 I will explain how PP can account for this Janus-faced capacity.

In sum, missing from the literature on dynamical approaches to emotion is an explanation of the formation of habitual emotional responses grounded in the repetition of emotional episodes of a matching type. I now expand on this proposal, focusing on the case of anger. Section 3 proposes that emotional episodes of anger are identifiable with the operation of a feedback loop in which physiological arousal biases attention toward threatening stimuli, in turn encouraging biased appraisals and ruminative processes that feed back to maintain heightened arousal. This mechanism leads the organism to appraise incoming stimuli in a biased fashion that favours the attribution of hostile intent. In Section 4 I propose that it is repeated instances of emotional episodes of the same type (in our case anger) that over time

[1] Whether Hebbian learning and PP are frameworks that can be integrated for explaining learning is an important question, but one that lies beyond the bounds of what I hope to accomplish here. Friston (2010) states that a gradient descent on free energy (changing synaptic connection strength to reduce free energy) is formally identical to Hebbian plasticity. Translated into a PP framework, Hebbian learning states that when presynaptic predictions and postsynaptic prediction errors are highly correlated, connection strength increases, so that predictions are able to suppress prediction errors more efficiently. Despite this formal equivalence, Hebbian learning and PP are distinct inasmuch as the former describes a recapitulation process through which learning occurs, while PP describes a representational or inferential process in which prediction is adjusted based on prediction error (Sumner et al, 2020).

produce an emotional disposition matching the repeated emotion type (producing in our case dispositional anger).[2]

3 Biased appraisals

A feedback loop is a system where outputs at time $t-1$ are routed back as system inputs at t. A positive feedback loop features outputs routed back as inputs that tend to amplify the strength of the incoming signal, or the level of perturbation of the system. (A simple example is a bank account with compound interest. Deposits into the account increase the total balance, and compound interest amplifies this effect.) Figure 1 illustrates a feedback loop, partly constitutive[3] of an emotional episode of anger, that involves a positive feedback loop in which physiological arousal functions to bias perception, which in turn biases appraisal, which in turn sustains or increases arousal. Figure 1 shows a stimulus from the environment at time t_1 initiating changes in the values taken by component measures of emotion, which causally interact with each other (arrows in black). These component interactions influence how the environmental stimulus is perceived by the system at time t_2, and the environmental stimulus in turn exerts further changes on the component parts (arrows in grey). In this way, changes in the components effect how the environmental stimulus is perceived in future moments of time. I first outline the empirical evidence supporting this picture and then use a familiar example – road rage – to give it some colour.

There is evidence supporting the view that what we perceive is modulated by what we attend to. The focus of attention changes frequently, and is influenced not only in a 'top-down' manner (i.e., forming the intention to pay attention to the words on a page), but also through pre-conscious directing of attention through 'bottom-up' influences. One such bottom-up influence is moment-by-moment changes in physiological arousal (Todd and Anderson, 2013; Beck and Kastner, 2009; Mather and Sutherland, 2011; Desimone and Duncan, 1995). One well-supported proposal for how this occurs is that arousal biases competition among stimulus presentations (Mather and Sutherland, 2011; Mather et al, 2016). These presentations initially compete in the sensory cortex for processing priority, with salient stimuli initially dominating, and thus initially attracting attention. Salient stimuli win the initial competition for selective attention when they are perceptually conspicuous or goal-relevant (Beck and Kastner, 2009; Desimone and Duncan, 1995). In this way, arousal shapes the attentional profile, increasing the attention paid to salient stimuli and inhibiting further the attention paid to non-salient stimuli.

[2] Adopting a dynamical approach to emotion here perhaps raises a broader question about the relationship between DST and PP. DST is frequently linked to approaches (e.g., Hohwy, 2016; Ward et al, 2017) which seek to understand cognition primarily in terms of embodied agent and environment dynamics. This is because DST provides an apparatus for describing the unfolding operations of complex systems composed of multiple closely interacting parts, in this way providing a tool for describing the evolving states of a system as it navigates its environment over time. But actually, DST itself is simply a tool to study temporal dynamics, i.e., differential equations are used to describe the ways in which the system can transition from one point on the state space to another (as opposed to the various concepts and theories DST is frequently related to, such as self-organisation), and it is this theory-neutral and narrower characteristion of DST I draw on here. My focus here will include a methodological suggestion about using a dynamical conception of emotion to get clearer about the nature of prediction and precision weighting, but for a broader reflection on the relationship between ecological psychology, embodied dynamics and PP, see Bruineberg and Rietveld (2014).

[3] I say 'partly' here because there are other subsystem processes that also partly constitute emotion, such as facial expression and action tendencies. A broader and more detailed account would incorporate these processes, but I present here a deliberately simplified model focused more narrowly on evaluative perception.

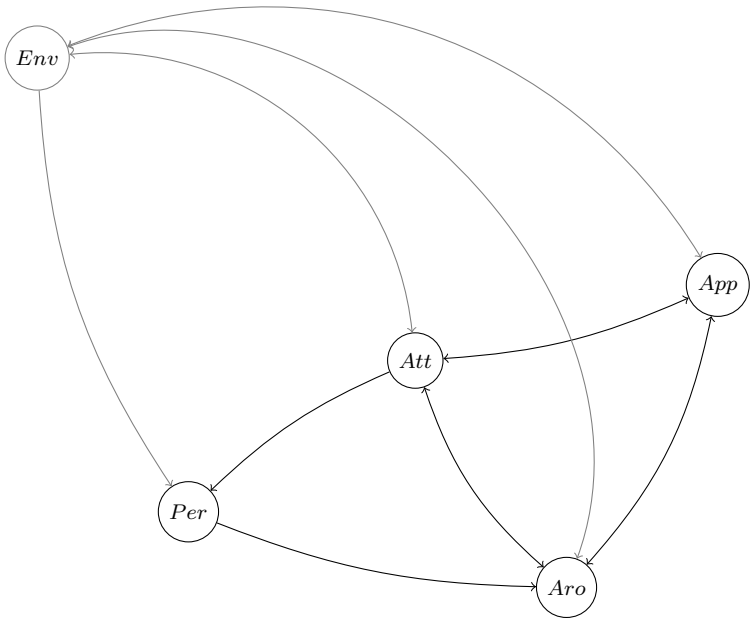

Fig. 1: Environmental input driving component changes across time. Component variables in this example include perception, arousal, attention, and appraisal. These components causally influence each other (black arrows). They also influence how the environmental stimulus itself is perceived, which in turn exerts further influence on the components themselves (grey arrows). Unidirectional arrows indicate a one-way relationship, bidirectional arrows indicate a two-way relationship.

At the neural level, attentional filtering is facilitated by arousal-induced norepinephrine (NE) and glutamate release that biases perception and memory in favor of salient, high priority representations at the expense of lower priority representations. This occurs via glutamate-modulated NE 'hot spots' at the site of prioritised representations. This excitatory effect contrasts with widespread NE suppression of weaker representations via lateral and auto-inhibitory processes. At a broader scale, hot spots increase oscillatory synchronisation across neural ensembles transmitting high priority information, while key brain structures preferentially route such information through large-scale functional brain networks (see Mather et al, 2016).

The neurophysiology stimulating attentional filtering maps closely to the physiological arousal associated with anger. Anger is an emotion that comes under the broader category of fight or flight excitement states, all of which share, at a course grain, a similar visceral pattern. This pattern, which has been researched since Cannon (1929), includes changes in blood pressure, heart rate, skin conductance, muscle potentiation, respiration rate, and hand and face temperature. Further sub-patterns can be distinguished among fight or flight states. Neurochemically, anger arousal includes increased phasic NE, with increased glutamate involved in anger modulation (Lara and Akiskal, 2006).

Appraisal is encouraged by this process. In the affective science literature, appraisal refers to that component of emotion through which an organism evaluates a stimulus in the environment in relation to its goals and purposes. This evaluation includes both an assessment of what is occurring, as well as a prescription of what ought to be done about it (appraisals are 'pushmi-pullyu' representations in the sense described by Millikan, 1995). There is consensus among appraisal theorists that anger involves either the appraisal that an event is goal-incongruent and must be remedied, or that it is goal-incongruent and that some third-party is blameworthy and deserving of retaliation (Silvia and Warburton, 2006; Kuppens et al, 2003). There is also consensus that appraisal can be implicit and more or less automatic (when a sharp poke in the back prompts your anger well before you spin around to determine the identity and intention of your assailant)[4] or explicit and accessible to conscious deliberation (when your assailant apologises profusely and you are left to judge whether their smile is more like a genuine attempt at appeasement or a smirk). This distinction was proposed by the 'mother' of appraisal theory, Magda Arnold. Arnold (1960) described, for instance, one variety of appraisal as 'direct, immediate, intuitive' (Arnold, 1960, p. 172) and 'hidden from introspection' (Arnold, 1960, p. 177), citing the example of a person moving to avoid someone who moves to stab their finger. In describing the second type, Arnold writes that 'intuitive appraisal is often supplemented or corrected by later reflection. When this happens, the emotion changes with the new intuitive estimate which follows the corrective judgment' (p. 175). This distinction between implicit and explicit appraisal is also generally accepted by contemporary appraisal theorists and is helped along by a tendency to define appraisal by function rather than underlying mechanism. Moors et al (2013) for instance, note that contemporary theorists claim that appraisal does not consist primarily of 'abstract cognitive principles'. The idea is that appraisal will frequently be automatic, and can comprise representations that are 'perceptual and/or embodied', with action tendencies that obviously manifest the evaluative content of the implicit appraisal. At other times, appraisal will be non-automatic and rule-based, operating on 'symbolic representations' (Moors et al, 2013, p. 121). These explicit appraisals are the type that present themselves to conscious reasoning.

Multiple lines of evidence suggest covariation and reciprocal influence between appraisal and arousal. First, as was noted earlier, the emotional concordance data shows a covariance pattern between multiple emotion components, two of which are appraisal and physiological arousal. Second, it has been proposed that increased positive functional connectivity with the thalamus during angry rumination reflects a pattern of reciprocal influence between the former and arousal. The idea here is that the executive functions typically supported by the prefrontal cortex (e.g., planning revenge) covary with a sense of heightened arousal (Denson, 2013).

Influence between arousal and appraisal is reciprocal. First, there is growing recognition that arousal (or perhaps perception thereof) likely motivates and influences reasoning processes (Haidt, 2013) This evidence comes from psychological studies showing that a person's 'gut reaction' to a vignette can drive their reasoning processes, and indeed in a biased way: toward a reasoned assessment of the situation that conforms to the initial feeling the vignette provoked (Haidt, 2001). It may also independently increase angry rumination by increasing the likelihood of the agent choosing hypotheses consistent with their interoceptive awareness of their bodily state. In the other direction, the perception of physiological activity, otherwise known as interoceptive awareness, can heighten the intensity of emotional arousal (Dunn et al, 2010). Consistent with the role of the insula in interoceptive awareness Denson et al (2009) have found, for instance, that right anterior insula activation is positively correlated

[4] This example is from Griffiths (2003).

with self-reported state angry rumination. Thalamus activation is also important due to its role in emotional processing, emotion experience, and emotional control (Marchand, 2010). So together, increased activation in the right anterior insula and thalamus during rumination may heighten the experience of emotional arousal (see also Denson, 2013).

We can think about this heuristically. To do so, let's use the same example discussed by Lewis (2005, 175) involving a driver called 'Mr. Smart', who experiences road rage when cut off suddenly on the highway by a speeding driver in a expensive-looking sports car. Say that at t_1 Smart initially perceives the rapidly vanishing distance between his car and the one changing lanes in front of him as salient (Figure 1, perception), which stimulates arousal (Figure 1, arousal) and prompts his immediately slamming the brakes and swearing at the driver in front. On an input-output model, this filtered perceptual data is the 'input' for an evaluative process culminating in what is initially an implicit appraisal of the stimulus as a life-threatening goal obstruction (e.g., Lazarus, 1991).

At time t_2 the increase in Smart's arousal features as input into the system, and serves to decrease his awareness of other features of his environment (the half-read newspaper on his lap, where the next freeway exit is, etc.) and increase his awareness of the object that jeopardised his safety a moment earlier. At t_3, attentional filtering (Figure 1, attention) leads to Smart perceiving visual stimuli relating to the car in front of him and its driver, rather than the newspaper he was focusing on a moment ago. In particular, the facial expression of the driver in front (somewhere between startle and hostility) appears salient and dominates Smart's perceptual hierarchy (for attentional bias toward facial expressions, see Calvo and Nummenmaa, 2008). Since Smart remains fixated on the driver's facial expression, he fails to notice other crucial details of the scene – one being the obviously injured person in the passenger seat of the car in front.

Failing to notice the injured passenger, Smart's initial appraisal of goal-obstruction now gets a little more personal, as he fixates on the potentially hostile expression of the driver in front and the flashy low aero-wedge on the bright red sports car. He now makes a more sophisticated evaluation, namely that he has not only had his goals (of getting to work on time, of being safe on the highway) obstructed, he has had these goals obstructed by a show-off who cut him off on purpose, and deserves retaliation. Smart verbalises this appraisal as best he can, e.g., by shouting at the driver, blasting his horn, and attenuating the counter-attack with a number of hastily-chosen expletives. This movement from an implicit, initial appraisal of goal obstruction and toward an explicit appraisal of other-blame is consistent with the well-known claim of Frijda (1993) that angry people look for someone to blame even if an appraisal of blame did not initially cause the anger.

Smart's developing appraisal (Figure 1, appraisal) now feeds back into the system, stimulating further physiological arousal and attentional filtering. The product is a feedback loop sustained by what seems to be a biased interpretation of the situation at hand, and continuing arousal that filters perception to encourage the development of further appraisals consistent with the thematic content of anger (i.e., the presence of a third party with hostile intentions deserving of blame or retaliation).

4 Habitually-biased appraisals

So far a dynamical model has been used to describe how a biased appraisal of a stimulus can be produced as part of a single emotional episode of anger, and how this biased appraisal can trigger a feedback loop sustaining the emotional episode. Now, if the situation for a person with HAB were analogous to a Foucault pendulum, such that an initial impetus set off some

feedback loop, and the system continued in a perpetual state of activation, our explanation would be complete. But a typical case of HAB will involve dispositional rather than perpetual anger: anger will be an attractor region on that individual's emotional state space, perhaps the strongest, but not the only one. The explanatory lacuna here comes in explaining how an HAB individual can be predisposed toward anger when the feedback loop associated with an emotional episode is not active. We need to explain not just the tendency for anger to continue once triggered, but the tendency for the system to move towards anger over other possible options on the state space (with at least greater likelihood than would be expected for an individual who did not have HAB).

Let me expand on this point by returning to Smart and considering two ways his emotional trajectory could progress. There are at least two types of intervention on Smart's emotional system that could work to sustain the feedback loop. The first is continued stimulus from the environment. Figure 2 illustrates a case in which Smart's anger subsides relatively quickly. t_1 represents an initial state of relative neutrality, where the relevant components involved in emotion are at their baseline (see the nodes in the middle of the vertical lines). Input from the environment at time t_1 causes component changes at time t_2. Through this initial trigger in the road rage case, Smart becomes primed to focus on aspects of the external environment that confirm his assessment of the speeding driver's blameworthiness. The driver's embodied reactivity to the event stays stable between time t_2 and time t_3, and this represents the slower bodily dynamics of the system (It is known that physiological arousal lingers after the initial stimulus, in some experimental settings up to 3,000 milliseconds assuming no further intervention on the system, see Mather and Sutherland, 2011). At time t_3 there is new input from the environment: the other driver turns around and stares at Smart, with a look of startle that seems ambiguous between fear and anger. Smart, affectively primed by his arousal, fixates his attention narrowly on the driver's facial expression and fails to attend to other features of the environment, including our injured passenger. The driver in front continues to embroider the highway, moving left and right between lanes, and Smart keeps driving, at a much less impressive speed, in the same lane. At t_5 the sports car driver's lips curl into a defensive snarl, and it is at this point that Smart starts yelling, then at t_7 the speeding driver yells back, and the shouting match evolves from there in a predictable fashion.

But negative rumination on the event on the highway could work just as well as external input in sustaining Smart's anger: The sports car driver speeds away, but Smart ruminates on the event on the highway all the way to work. Since we are assuming reciprocal influence between rumination and heightened arousal (discussed above), Smart remains in a state of heightened arousal. Consequently, he remains primed to filter later events in a manner consistent with appraisals involving a projection of hostile intent. So, for example, Smart makes it to work, starts chairing a meeting, but is quickly interrupted by a colleague. The colleague might have been motivated by pure enthusiasm for the topic, or perhaps had not heard Smart begin to speak, but Smart, already primed to filter his attention in a manner consistent with the appraisal of other-blame, experiences the interruption as a deliberate affront. This perception triggers Smart from a baseline of residual irritation toward a renewed 'hot' anger, and this renewed surge of anger sustains heightened arousal, in turn encouraging further rumination consistent with goal-obstruction and other-blame.

But what if neither of these two things occurs? At t_5 the speeding driver's face turns from part startle, part hostility to a look of genuine concern and regret, and he then looks away, cautiously restarts his engine, and drives off. Smart starts driving again too and encounters no further attempts at one-upmanship on the highway. He also starts deliberately employing emotional regulation techniques to calm himself (e.g., he decides not to ruminate on what

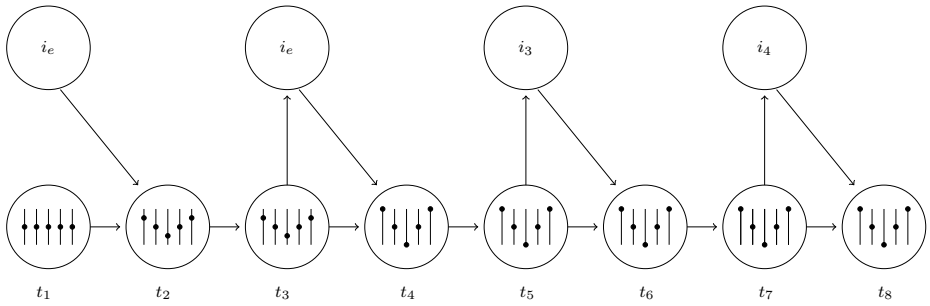

Fig. 2: Environmental input at time t_1 drives component synchronisation at time t_2. Embodied reactivity lingers between t_2 and t_3 due to slower bodily dynamics. Environmental input at time t_3 drives further component synchronisation at time t_4 and is mediated by attentional filtering of this input. Embodied reactivity again lingers between t_4 and t_5 and the process repeats itself, creating a positive feedback loop. Component synchronisation is measured via changes in the values taken by representative variables, each of which is represented as a node that can move up or down a vertical axis as its value changes.

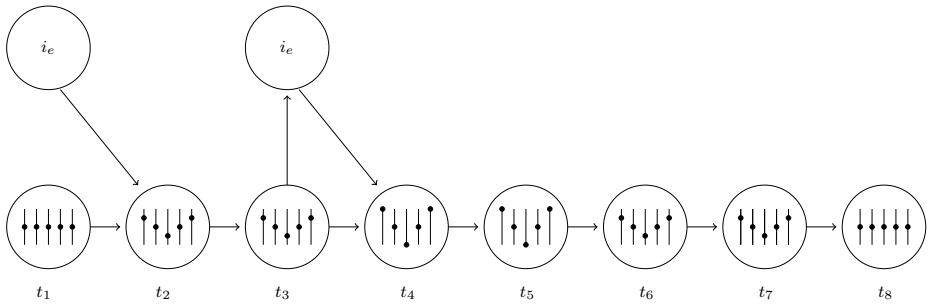

Fig. 3: Environmental input at time t_1 drives component synchronisation at time t_2. Embodied reactivity lingers between t_2 and t_3 due to slower bodily dynamics. Environmental input at time t_3 drives further component synchronisation at time t_4 and is mediated by attentional filtering of this input. Lacking further perturbation, the system slower returns to baseline between t_5 and t_8. Component synchronisation is measured via changes in the values taken by representative variables, each of which is represented as a node that can move up or down a vertical axis as its value changes.

just happened, practices mindful breathing, turns on the radio for distraction, etc.). The remainder of the journey is uneventful, and so, between t_5 and t_8, Smart's anger rescinds and he returns to a more or less neutral, 'baseline' state (see Figure 3).

So then, if the state space representing the emotional possibilities for someone with HAB includes more-or-less neutral regions, we need to be able to explain why the system remains prone to making appraisals consistent with HAB, *even once* it has returned to a more-or-less neutral state. And this is where drawing on aspects of a PP framework may assist. My proposal in the remainder of this section is that the HAB person's generative model is 'bent' over time by biased appraisals. This eventually produces a mature system that will reliably produce biased appraisals that are relatively immune to modulating influence from

the outside, even when the feedback loop described in the previous section is not running continuously.[5]

An agent's generative model M aims to capture the statistical structure of some set of observed inputs by tracking the causal matrix responsible for that structure (Clark, 2013). Recall that a PP framework proposes that the brain and nervous system interprets the meaning of events in the world by correctly anticipating (predicting and adjusting to) incoming sensations. It also proposes that the brain and nervous system is constantly assembling populations of predictions, each of which has some probability of being the best fit to the current circumstances. These probabilities are Bayesian priors, and are implemented in the brain as distributed patterns of activity across certain populations of neurons. For every hypothesis the organism could generate in the present about what is currently occurring, there is a prior probability that that particular hypothesis is true.

Prior probabilities are based on the system's life history. Someone with many previous experiences of unhelpful and dismissive waiters will assign a higher prior probability to the hypothesis that the waiter currently approaching the table is an unhelpful one. Incoming sensory evidence – prediction error – will help select from or modify the distribution of predictions, because certain predictions will better fit the sensory array (i.e., will have stronger priors), with the end result that incoming sensory events will be categorised as being similar to some set of past experiences. The combination of sensory input and the prior probabilities of predictions is used to reduce the size of the population of generated predictions and thereby settle on a single hypothesis. As a result, a hypothesis could fit the incoming sensory input extremely well, but its prior probability could be so low that it is ignored. Conversely, a hypothesis could have such a high prior probability that, even though it doesn't fit the sensory input well at all, it is selected.

Now we have already seen that there is plenty of support for the idea, discussed in Section 2, that affect is centrally involved in prediction generation. But there also seems to be a plausible case for identifying the appraisals that partly constitute emotional episodes with predictions about the world — ones that feed into M as they are iteratively generated over the course of the organism's interactions with its environment over time. For one, we have seen that appraisals broadly describe events in the world in both descriptive and imperatival form (as varieties of 'pushmi-pullyu' representations). Anger, for instance, involves an evaluative perception of goal obstruction caused by an actor with hostile intention who is worthy of blame (or some other form of retaliation). Higher-level predictions share this two-fold structure since they too are affordance-laden, containing both sensory and motor predictions relating to the world (Clark, 2016; Friston et al, 2012).[6] For two, the predictions that feature in M specify the probability that a certain external state, sensory input, and internal state occur together (Friston, 2013), and we also see this schemata in the emotion case. For example, we saw that the evaluative perception associated with anger is also responsive to internal input, so that the evaluation that a blameworthy actor with hostile intent is present (e.g., the speeding driver) is hypothesised to be the cause of both the internal state of the body (e.g., heightened arousal)[7] and sensory input (e.g., the visual input relating to the speeding

[5] I am grateful to an anonymous reviewer whose considered feedback helped me to see why a dynamical model may, in and of itself, be insufficient to explain HAB, and for their encouragement to explore how emotion might play a role in creating, updating, and sustaining long-term predictions in greater depth.

[6] This is not to exclude the possibility that lower-level predictions also share this two-fold structure. I only emphasise higher-level predictions here because I have been focusing on appraisals that involve the thematic content of blame attribution.

[7] For details of this subcortical processes involved in this communicative process between the nervous system and brain see Miller and Clark (2018).

driver described above). There is also recent work (Barrett, 2017) that presents a structural model of the corticocortical connections that may underwrite emotional episodes as types of 'concept' (Barsalou, 1983, 2003), namely the type involving categorising some situation and set of bodily feelings as instances of some emotion category type (of fear, or happiness, or anger), and manifest as a prediction or distribution of predictions matching the thematic appraisal content of that emotion category type (see also Wilkinson et al, 2019). Of course, the plausibility of casting emotional appraisals as a variety of prediction does not on its own motivate the use of PP as a framework for thinking about emotion. But so doing can, I think, help to make sense of the development of emotional dispositions.

I now want to combine the hypothesis that biased appraisals will bend the generative model over time with the proposal that the feedback loop facilitating such appraisals is a potential mechanism for the implementation of precision weighting.[8] Precision weighting refers to how much epistemic weight is given to top-down predictions versus prediction error (sensory input). Expected precision is thought to be altered by modulating the synaptic gain (postsynaptic responsiveness) of prediction error units (Feldman and Friston, 2010). For example, if low precision weighting is attached to prediction error at a given moment, a prediction might be settled on that fits the incoming sense data very poorly. And indeed this is what happens in the case of the individual with HAB: they see threats in their environment largely because they are *expecting* to see them. This suggests a role for arousal-induced attentional filtering as a method for modulation of precision weighting in the brain and nervous system.[9] This proposal is also supported by the the work of Miller and Clark (2018) who outline a neuroanatomical model suggesting that attention increases selectivity for prediction error.[10]

Let's sum up. We have seen, so far, that a dynamical conception of emotion explains the inbuilt tendency for emotional episodes of anger to produce biased appraisals, and for these episodes to be sustainable with fairly minimal intervention. But to explain how HAB can arise from a neutral place on the state space we need to see each appraisal formed during an emotional episode as a prediction that feeds into the agent's generative model of what the environment affords. Each new biased appraisal (one-upmanship on the highway, a demeaning interruption during a meeting) bends the generative model, changing the prior probability the system will in future attach to the hypothesis that some novel situation involves an actor with hostile intentions. We now have an account of a system whose initial conditions in early development, combined with an inbuilt mechanism for emotional modulation of

[8] I am grateful to an anonymous reviewer who assisted in the development of this line of argument.

[9] See Ransom et al (2020) for further discussion of how the broad hypothesis that affect-biased attention modulates precision weighting modulation could be precissified and further investigated. It is noteworthy that Ransom et al (2020) discuss different varieties of affect-biased attention, including those involving stimuli that strike the perceiver as salient even when the precision weighting of prediction error is low. Representative examples include habitual attentional orientation towards a fence where a ferocious Doberman was seen only once, or to situational features that resemble a single past traumatic event (e.g., so-called Type 1 trauma, see Terr, 1991). I have focused here on a variety of affect-biased attention that arises from repeated events of the same type: ones that generate, over time, a habitual attentional orientation. The apparatus of a generative model continuously updated by 'biased' appraisals of the type here described seems well-suited to explaining the phenomenon of HAB (and it is possible the same general schematic might be useful in explaining PTSD arising through so-called Type 2 trauma). Further work would be required, however, to explain how this apparatus might be further developed to incorporate 'Type 1-like' cases of habitual attentional orientation. For further discussion of this issue, see Ransom et al (2020), and for an interesting PP-based proposal applied to both Type 1 and Type 2 trauma in the context of PTSD see Wilkinson et al (2017).

[10] Miller and Clark (2018) suggest that emotion modulates precision via the activity of the pulvinar complex, but I do not think that this claim need be understood as limiting the essential dynamics involved in this process to the brain. I expand on this point below in footnote 13.

precision weighting, gradually produce a generative model in which predictions involving actors with hostile intentions feature disproportionately highly. Importantly, this can be so even when the 'early training ground' of the system is one featuring social environments in which genuine hostility is present (thus the account is responsive to what was earlier described as the Janus-faced functionality of emotion, see Section 2.) In this way PP can explain how emotions can be responsive to both the present input and a past learning history across an interval of time corresponding to an emotional episode. The end product is a system that is, in its mature state, relatively immune to sensory input that might seek to disconfirm the general view of a world made up of actors with hostile intentions, due both to the generative model's according such a perspective a disproportionately high likelihood, combined with the tendency of emotional episodes of anger to confirm such expectations due to the modulation of precision weighting.

For now, it remains to connect this account to the original problematic: explaining the correlation between early psychosocial stress and the development of organic disease phenotypes in midlife. Before proceeding to this, I briefly mention a distinctive methodological advantage of bringing a dynamical approach to emotion into dialogue with a PP framework. As we've seen, identifying emotions with episodes of synchronisation provides temporal markers that make it possible to narrow down – to within milliseconds – when we can expect biased predictions of the type I've described to occur.[11] This in turn makes it possible to compare the temporal dynamics operative during these synchronisation windows against the dynamics observed when the system is in a more neutral state, to test the hypothesis that certain emotion types modulate precision weighting to favour prediction (for example, through comparing how the gain on select populations of prediction error units is altered during temporal windows featuring such episodes against those of relative emotion-neutrality). This would increase our understanding of the temporal dynamics already known to be associated with precision weighting modulation, in turn providing further support for the proposal that emotion, attention, and prediction are intricately intertwined.[12] Over a longer timescale, the iterative structure of the history of the system's interactions with its environment could be mapped to gradual changes in candidate implementation sites for an embodied generative model (one possibility being the development of perpetually heightened measures of arousal, such as blood pressure and heart rate variability).[13] The identification of emotions as dynamical episodes can be used in this way to test hypotheses about how (distinctively emotional)

[11] For a discussion of the nature of data available and methods used to identify windows of synchronisation, see Bulteel et al (2014); Hollenstein and Lanteigne (2014).

[12] These dynamics are explored using EEG and fMRI to measure changes in oscillatory band frequency and event-related potential, see, for instance Smout et al (2019).

[13] I have here attempted to motivate the role the generative model can play in explaining the formation of emotional dispositions. But this is just one step towards a fuller picture of how the model may be implemented in the brain and nervous system, or perhaps more broadly as embodied and enacted in the whole organism itself as it navigates its environment (e.g., Allen and Friston, 2018; Friston, 2013). I cannot delve in detail into this issue here, beyond noting that the proposal outlined here seems to be compatible with a 'radically embodied' PP (Friston, 2013; Allen and Friston, 2018). Such an approach sees the generative model of the world embodied in a web of neural connections of varying strengths, and causally coupled to the body, specifically its homeostatic needs and the environmental niche within which it has evolved. On this approach there is a sense in which homeostatic set points (e.g., blood pressure and blood glucose levels) partly constitute the organism's generative model, even though the neural connections forged through a particular organism's unique learning history also constitute this same model. This more extensive construal of the generative model is consonant with the suggestion made in Section 3 that slower bodily dynamics of arousal can lengthen the time interval during which the feedback loop constituting an emotional episode is operative. If the homeostatic set points relevant to such arousal are conceived as part of the generative model, we have a putative mechanism through which the former might influence the development of real-time emotional episodes, triggering attentional filtering and biased appraisals.

predictions are implemented in the brain and nervous system, in turn shedding light on how implementing this type of prediction may differ from implementation of predictions occurring during relatively 'non-emotional' periods,[14] while also providing a means to model how early life experiences shape mature phenotypic emotional expression.

5 Organic disease with a top-down cause

We can now return to the relationship between early psychosocial stress and later susceptibility to disease in mid-life. The intention was to explain how certain top-down processes (namely, subjective perceptions and evaluations of the social and material environment) were able to exert a profound influence on the lower-level physiological function and structure of the body. For simplicity, I focused on one simple and easily measured physiological marker: persistent elevated blood pressure (hypertension), and its connection with habitual anger.

To be clear, the puzzle here is not so much explaining a temporary elevation of blood pressure concomitant with a surge of anger. We already know that elevated blood pressure is part of the physiological signature of anger: This physiological signature has aspects common to all fight or flight responses, but there are features distinguishing an anger response in particular. As anger emerges, blood pressure heightens and remains elevated so long as the episode of anger persists (Garfinkel et al, 2016). For example, one of the very first studies to differentiate the physiological signatures of fear and anger (both of which form a broadly fight or flight pattern of response) found that anger produces greater changes in diastolic blood pressure rise, heart rate fall, rise in skin conductance, and muscle potential increases (Ax, 1953). And we also know that anger can reliably be induced or sustained through biased inferences of the type associated with HAB (itself one component of the synchronisation response associated with anger).

The actual puzzle, then, is explaining why blood pressure might remain elevated in an individual throughout the lifespan, long after that person is removed from the environment that originally caused the initial episodes of arousal-inducing anger. The answer I have explored here is that the individual in question suffers chronically elevated blood pressure because they feel habitual anger. To explain how dispositional anger could develop gradually over time in response to early psychosocial stress, I suggested that we see an emotional disposition toward anger as simply sustained anger whose presence is relatively resistant to outside influence. I identified emotional episodes of anger with a feedback loop involving perception, attentional filtering, and arousal, and described how this loop generated biased appraisals and ruminative processes that fed back into the loop to sustain it.

I noted that the feedback loop can explain the generation of biased appraisals only whilst it is operative. Consequently, an additional mechanism is needed to explain how a tendency to form biased appraisals could develop and become entrenched over time, one that does not depend on the feedback loop running consistently, else HAB would need to be characterised as not a habit or tendency toward anger, but as – rather unrealistically – a perpetually present feature of the system. I then proposed that the appraisals involving HAB, generated by the feedback loop, come to bend the agent's generative model over time. The feedback loop can be initially set off by input that actually does involve hostile intent (such as abuse or violence). The trick of the feedback loop is that this leads to lingering arousal that tips the system toward attentional filtering that encourages interpreting future stimuli – including

[14] For a discussion of how to distinguish 'emotional' versus 'non-emotional' periods of operation of a system, see Colombetti (2014).

neutral stimuli – in a manner consistent with the appraisal themes associated with anger. I proposed that this tendency could be seen as a mechanism through which precision weighting favouring prediction over prediction error is implemented in the system. The development of dispositional anger, then, is explained via two key proposals. The first is that HAB-involving appraisals are encouraged by a positive feedback loop in which physiological arousal biases attention toward threatening stimuli, in turn encouraging biased appraisals and ruminative processes that feed back to maintain heightened arousal. The second is that HAB-involving appraisals come to bend the agent's generative model over time. Each new HAB-involving appraisal affects both the generative model as well as precision weighting over the short term, while the generative model – which is coming to represent an increasingly hostile and threatening world – informs the generation and selection of subsequent predictions.

The product is an account that connects the agent's subjective perception of their environment to changes in physiological function manifest both in real time (elevated blood pressure) and developmental time (persistent elevated blood pressure): It is known that persistent elevated blood pressure gradually causes the coronary arteries serving the heart to slowly become narrowed from a buildup of fat, cholesterol and other substances that together are called plaque (Cohen and Hasselbring, 2007). Continued over a long enough time, hypertension becomes a risk factor for stroke, coronary heart disease, heart failure, and end-stage renal disease (Muntner et al, 2014). There are, of course, likely to be multiple causes in the case of any given individual manifesting the phenotype under scrutiny. Some of these factors may be more influential than others, and knowing which are may have a direct bearing not only on the understanding of disease aetiology, but also on which treatments are most appropriate for given individuals. We would expect, for example, that emotion will play an important explanatory role in cases in which the psychosocial environment plays a primary role. In these cases psychosocial intervention might be useful. However there will of course also be cases of hypertension for which the main causal difference-maker does not concern the psychosocial environment, e.g., hypertension caused by HIV infection or sickle cell disease (Galie et al, 2004). These cases would, of course, require an entirely different sort of treatment.

Stepping back, the development of HAB can be seen as an instance of developmental mismatch, which arises when the organism in early life uses the mechanisms of adaptive developmental plasticity to develop a phenotype suitable for one environment, but is later exposed to a different environment from that to which it has adapted (Bateson et al, 2004; Gluckman et al, 2019; Gluckman and Hanson, 2007).[15] If HAB developed in a person placed in a genuinely hostile environment (e.g., an abusive family, a period of war or civil unrest) and they lived out their entire life under these conditions, HAB would be highly adaptive, enabling quick and reflex-like responses to aggression. But in a context where the organism is free to choose in maturity a relatively benign social and political environment, or finds themself in one through happenstance, the automated defensive behaviours prompted by HAB no longer serve an adaptive function. Regardless of whether it is adaptive for the organism's current environmental context, the physical costs of this emotional disposition (e.g., premature mortality due to cardiovascular disease) will remain. Placed in a sufficiently hostile setting, premature morality is perhaps a reasonable trade-off for an expedient strategy to quell genuine aggression, but in a benign context it is not. (It is these messy realities that reveal the imperfect solutions evolution provides to solve problems of adaptation.)

[15] I am grateful to Paul Griffiths and members of the Theory and Method in Biosciences Lab, Charles Perkins Centre, University of Sydney for useful discussion of this idea.

6 Conclusion

For much of its history, emotion psychology favoured a simple input-output model in which environmental input triggered an appraisal ('This snake is dangerous'), in turn triggering an emotional response (sympathetic nervous system activation, running away). This approach, which received its canonical expression in the work of Ekman (1992, 2003) did allow life experience to alter the eliciting conditions of an emotion, e.g., so a history of encountering nurses giving painful injections could train the emotional system to immediately react with fear upon seeing a nurse. But explaining a mature emotional disposition demands more than a simple input-output correspondence structure (input: nurse, output: fear), since such dispositions consists of habitual patterns of response that seem to be stimulated by a tremendously broad set of conditions: consider a generalised anxiety that manifests in continued hypervigilance in many situations encountered in daily life (not just in hospitals) or, as in our case, a tendency towards attributing hostile intent to a wide range of actors whose intentions are utterly benign (a tendency whose source lies in early traumatic experiences, but whose current exercise extends far beyond the specific people and environments in which those early experiences occurred).

The role of emotional episodes in this explanation is as a means through which an organism's generative model is updated in iterative fashion over time in response to life experience, eventually producing a mature emotional disposition such as HAB, whose chief characteristic is insensitivity to incoming stimuli that would seek to disconfirm the evaluative perception of a world made up of blameworthy actors with hostile intentions. When we characterise emotions as 'dynamical' we pave the way for a definition identifying them – and the biased appraisals that partly constitute them – with transient synchronisation of measurable components that can be mapped to a temporal window. This definition grounds the conceptual proposal that early experiences shape later phenotypic emotional expression in a model that can be formalised and tested, through investigating the temporal dynamics of the cortical and subcortical processes at play during the temporal windows in which emotional episodes of particular category types (e.g., anger) occur.

In closing, I note one upshot of this account. If its (very rough) contours are correct, it gives us the outlines for a new sort of explanation of the developmental origins of certain health and disease outcomes, one that is not currently accommodated by a biomedical explanatory mode that focuses on monogenic, bottom-up and atemporal causation. The case of 'pathological prediction' I have described here is in fact just one particular type of top-down process that might connect psychological mechanisms (in particular, those relating to how we perceive and evaluate our social and material environments) to structural and functional changes in the human body (the placebo effect is another such top-down process, albeit one that occurs over a much shorter timeframe). This new sort of explanatory model represents a move beyond the idea of the 'root' cause of disease being monogenic and atemporal – as if it could occur in abstraction from the processual impact of the environment and the value we attribute to it as organisms. A consequence is that the idea that there is a 'mind-to-matter' relation, or, more specifically, that our subjective perception of situations we experience in our environments should have some influence on the structure and function of our bodies should no longer appear strange or inexplicable. Key here is the recognition that emotion is continually serving a coordinative function for the human organism. This function involves both a 'minded' (evaluative, perceptual) aspect and a bodily aspect. At the same time, emotion functions as a learning mechanism for an organism to minimise prediction errors relating to its perception of the environment and the action opportunities it affords. Understood in this way, the idea that our bodies might change over time in response

to our subjective (and perhaps idiosyncratic) perceptions of our environmental context is deeply unmysterious.

References

Allen M, Friston KJ (2018) From cognitivism to autopoiesis: towards a computational framework for the embodied mind. Synthese 195(6):2459–2482, DOI 10.1007/s11229-016-1288-5, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5972168/

Anderson KB, Graham LM (2007) Hostile attribution bias, Encyclopaedia of Social Psychology. SAGE Publications, Inc., California

Arnold MB (1960) Emotion and personality. Columbia University Press, New York, NY

Ax AF (1953) The physiological differentiation between fear and anger in humans. Psychosomatic Medicine 15(5):433–442

Badcock PB, Davey CG, Whittle S, Allen NB, Friston KJ (2017) The depressed brain: An evolutionary systems theory. Trends in Cognitive Sciences 21(3):182–194

Barrett LF (2017) The theory of constructed emotion: An active inference account of interoception and categorization. Social Cognitive and Affective Neuroscience 12(1):1–23

Barrett LF, Simmons WK (2015) Interoceptive predictions in the brain. Nature Reviews Neuroscience 16(7):419–429, DOI 10.1038/nrn3950

Barsalou LW (1983) Ad hoc categories. Memory & Cognition 11(3):211–227

Barsalou LW (2003) Situated simulation in the human conceptual system. Language and Cognitive Processes 18(5-6):513–562

Bateson P, Barker D, Clutton-Brock T, Deb D, D'Udine B, Foley RA, Gluckman P, Godfrey K, Kirkwood T, Lahr MM, McNamara J, Metcalfe NB, Monaghan P, Spencer HG, Sultan SE (2004) Developmental plasticity and human health. Nature 430(6998):419–421

Beck DM, Kastner S (2009) Top-down and bottom-up mechanisms in biasing competition in the human brain. Vision Research 49(10):1154–1165

Bruineberg J, Rietveld E (2014) Self-organization, free energy minimization, and optimal grip on a field of affordances. Frontiers in Human Neuroscience 8

Bulteel K, Ceulemans E, Thompson RJ, Waugh CE, Gotlib IH, Tuerlinckx F, Kuppens P (2014) DeCon: A tool to detect emotional concordance in multivariate time series data of emotional responding. Biological Psychology 98:29–42

Calvo MG, Nummenmaa L (2008) Detection of emotional faces: Salient physical features guide effective visual search. Journal of Experimental Psychology 137(3):471–494

Cannon WB (1929) Bodily changes in pain, hunger, fear and rage. Appleton, New York, NY

Carel H (2011) Phenomenology and its application in medicine. Theoretical Medicine and Bioethics 32(1):33–46

Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences 36(3):181–204

Clark A (2016) Surfing uncertainty: Prediction, action, and the embodied mind. Oxford University Press, London, England

Cohen B, Hasselbring B (2007) Coronary heart disease: A guide to diagnosis and treatment. Addicus Books, google-Books-ID: Ja9YAwAAQBAJ

Colombetti G (2014) The feeling body: Affective science meets the enactive mind. MIT Press, Cambridge, MA

Van de Cruys S (2017) Affective value in the predictive mind. In: Metzinger T, Weise W (eds) Philosophy and predictive processing: 24, MIND Group, Frankfurt, Germany

Denson TF (2013) The multiple systems model of angry rumination. Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc 17(2):103–123

Denson TF, Pedersen WC, Ronquillo J, Nandy AS (2009) The angry brain: Neural correlates of anger, angry rumination, and aggressive personality. Journal of Cognitive Neuroscience 21(4):734–744

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annual Review of Neuroscience 18(1):193–222

Dodge KA, Somberg DR (1987) Hostile attributional biases among aggressive boys are exacerbated under conditions of threats to the self. Child Development pp 213–224

Dolezsar CM, McGrath JJ, Herzig AJM, Miller SB (2014) Perceived racial discrimination and hypertension: A comprehensive systematic review. Health Psychology 33(1):20–34

Dunn BD, Galton HC, Morgan R, Evans D, Oliver C, Meyer M, Cusack R, Lawrence AD, Dalgleish T (2010) Listening to your heart. How interoception shapes emotion experience and intuitive decision making. Psychological Science 21(12):1835–1844

Ekman P (1973) Darwin and facial expression: A century of research in review. Academic Press, Cambridge, MA

Ekman P (1980) Biological and cultural contributions to body and facial movement in the expression of emotions. In: Rorty A (ed) Explaining emotions, University of California Press, Berkeley, CA, pp 73–102

Ekman P (1992) An argument for basic emotions. Cognition & Emotion 6(3-4):169–200

Ekman P (1999) Basic emotions. In: Dalgleish T, Power T (eds) The handbook of cognition and emotion, Sussex, UK: John Wiley and Sons, Ltd, pp 45–60

Ekman P (2003) Emotions revealed: Understanding faces and feelings. Weidenfeld & Nicholson, London

Ekman P, Cordaro D (2011) What is meant by calling emotions basic. Emotion Review 3(4):364–370

Ekman P, Friesen WV, Ellsworth P (1972) Emotion in the human face: Guidelines for research and an integration of findings. Pergamon Press, Elmsford, N.Y.

Feldman H, Friston K (2010) Attention, uncertainty, and free-energy. Frontiers in Human Neuroscience 4

Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, Koss MP, Marks JS (1998) Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The Adverse Childhood Experiences (ACE) study. American Journal of Preventive Medicine 14(4):245–258

Frijda NH (1993) The place of appraisal in emotion. Cognition & Emotion 7(3-4):357–387

Friston K (2013) Life as we know it. Journal of The Royal Society Interface 10(86):20130,475

Friston KJ, Shiner T, FitzGerald T, Galea JM, Adams R, Brown H, Dolan RJ, Moran R, Stephan KE, Bestmann S (2012) Dopamine, affordance and active inference. PLoS Computational Biology 8(1):e1002,327

Galie N, Torbicki A, Barst R, Dartevelle P, Haworth S, Higenbottam T, Olschewski H, Peacock A, Pietra G, Rubin LJ, others (2004) Guidelines on diagnosis and treatment of pulmonary arterial hypertension: The Task Force on Diagnosis and Treatment of Pulmonary Arterial Hypertension of the European Society of Cardiology. European Heart Journal 25(24):2243–2278

Garfinkel SN, Zorab E, Navaratnam N, Engels M, Mallorqui-Bague N, Minati L, Dowell NG, Brosschot JF, Thayer JF, Critchley HD (2016) Anger in brain and body: the neural and physiological perturbation of decision-making by emotion. Social Cognitive and Affective Neuroscience 11(1):150–158

Gluckman PD, Hanson MA (2007) Developmental plasticity and human disease: research directions. Journal of Internal Medicine 261(5):461–471

Gluckman PD, Hanson MA, Low FM (2019) Evolutionary and developmental mismatches are consequences of adaptive developmental plasticity in humans and have implications for later disease risk. Philosophical Transactions of the Royal Society B: Biological Sciences 374(1770):20180,109

Griffiths PE (1997) What emotions really are: The problem of psychological categories. University of Chicago Press, Chicago

Griffiths PE (2003) Emotions as natural and normative kinds. Philosophy of Science 71(5):901–911

Haidt J (2001) The emotional dog and its rational tale: A social intuitionist approach to moral judgment. Psychological Review 108(4):814–834

Haidt J (2013) The righteous mind: why good people are divided by politics and religion. Vintage, New York, NY

Herrald MM, Tomaka J (2002) Patterns of emotion-specific appraisal, coping, and cardiovascular reactivity during an ongoing emotional episode. Journal of Personality and Social Psychology 83(2):434

Hohwy J (2016) The self-evidencing brain. Noûs 50(2):259–285

Hollenstein T, Lanteigne D (2014) Models and methods of emotional concordance. Biological Psychology 98:1–5

Kim J (1999) Making sense of emergence. Philosophical Studies 95(1):3–36

Kim J (2006) Emergence: Core ideas and issues. Synthese 151(3):547–559

Kuppens P, Van Mechelen I, Smits DJM, De Boeck P (2003) The appraisal basis of anger: specificity, necessity and sufficiency of components. Emotion 3(3):254–269

Lara DR, Akiskal HS (2006) Toward an integrative model of the spectrum of mood, behavioral and personality disorders based on fear and anger traits: II. Implications for neurobiology, genetics and psychopharmacological treatment. Journal of Affective Disorders 94(1):89–103

Lazarus RS (1991) Cognition and motivation in emotion. American Psychologist 46(4):352

Lewis M (2017) Addiction and the Brain: Development, Not Disease. Neuroethics 10(1):7–18, DOI 10.1007/s12152-016-9293-4, URL https://doi.org/10.1007/s12152-016-9293-4

Lewis MD (2005) Bridging emotion theory and neurobiology through dynamic systems modeling. Behavioral and Brain Sciences 28(2):169–194

Lewis MD, Liu Zx (2011) Three time scales of neural self-organization underlying basic and nonbasic emotions. Emotion Review 3(4):416–423

Marchand WR (2010) Cortico-basal ganglia circuitry: A review of key research and implications for functional connectivity studies of mood and anxiety disorders. Brain Structure & Function 215(2):73–96, DOI 10.1007/s00429-010-0280-y

Mather M, Sutherland MR (2011) Arousal-biased competition in perception and memory. Perspectives on Psychological Science 6(2):114–133

Mather M, Clewett D, Sakaki M, Harley CW (2016) Norepinephrine ignites local hot spots of neuronal excitation: How arousal amplifies selectivity in perception and memory. The Behavioral and brain sciences 39:e200, DOI 10.1017/S0140525X15000667, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5830137/

Meuleman B (2015) Computational modeling of appraisal theory of emotion. PhD thesis, University of Geneva, URL https://archive-ouverte.unige.ch/unige:83638

Miller M, Clark A (2018) Happily entangled: prediction, emotion, and the embodied mind. Synthese 195(6):2559–2575

Millikan RG (1995) Pushmi-pullyu representations. Philosophical Perspectives 9:185–200
Moors A, Ellsworth PC, Scherer KR, Frijda NH (2013) Appraisal theories of emotion: State of the art and future development. Emotion Review 5(2):119–124
Morag T (2016) Emotion, imagination, and the limits of reason. Routledge, Abingdon, Oxon; New York, NY
Muntner P, Davis BR, Cushman WC, Bangalore S, Calhoun DA, Pressel SL, Black HR, Kostis JB, Probstfield JL, Whelton PK, others (2014) Treatment-resistant hypertension and the incidence of cardiovascular disease and end-stage renal disease. Hypertension pp 1012–1021
Ongaro G, Ward D (2017) An enactive account of placebo effects. Biology & Philosophy 32(4):507–533
Pessoa L (2013) The cognitive-emotional brain: From interactions to integration. The MIT Press, Cambridge, Massachusetts
Ransom M, Fazelpour S, Markovic J, Kryklywy J, Thompson ET, Todd RM (2020) Affect-biased attention and predictive processing. Cognition 203:104,370, DOI 10.1016/j.cognition.2020.104370, URL https://linkinghub.elsevier.com/retrieve/pii/S001002772030189X
Scherer KR (2009a) The dynamic architecture of emotion: Evidence for the component process model. Cognition and Emotion 23(7):1307–1351
Scherer KR (2009b) Emotions are emergent processes: They require a dynamic computational architecture. Philosophical Transactions of the Royal Society B: Biological Sciences
Seth AK (2013) Interoceptive inference, emotion, and the embodied self. Trends in Cognitive Sciences 17(11):565–573
Silvia PJ, Warburton JB (2006) Positive and negative affect: Bridging state and traits. In: Thomas JC, Segal DL (eds) Comprehensive Handbook of Personality and Psychopathology , Personality and Everyday Functioning, John Wiley & Sons
Slavich GM, Cole SW (2013) The emerging field of human social genomics. Clinical Psychological Science: A Journal of the Association for Psychological Science 1(3):331–348
Smout CA, Tang MF, Garrido MI, Mattingley JB (2019) Attention promotes the neural encoding of prediction errors. PLoS Biology 17(2), pLOS Biology 17(2): e
Stemmler G, Aue T, Wacker J (2007) Anger and fear: Separable effects of emotion and motivational direction on somatovisceral responses. International Journal of Psychophysiology 66(2):141–153
Sumner RL, Spriggs MJ, Muthukumaraswamy SD, Kirk IJ (2020) The role of Hebbian learning in human perception: a methodological and theoretical review of the human Visual Long-Term Potentiation paradigm. Neuroscience & Biobehavioral Reviews 115:220–237, DOI 10.1016/j.neubiorev.2020.03.013, URL http://www.sciencedirect.com/science/article/pii/S0149763419310942
Svenaeus F (2013) Naturalistic and phenomenological theories of health: Distinctions and connections. Royal Institute of Philosophy Supplement 72:221–238
Terr L (1991) Childhood traumas: An overview and outline. American Journal of Psychiatry 148:10–20
Todd RM, Anderson AK (2013) Salience, state, and expression: the influence of specific aspects of emotion on attention and perception. The Oxford Handbook of Cognitive Neuroscience 2:11–31
Ward D, Silverman D, Villalobos M (2017) Introduction: The varieties of enactivism. Topoi 36(3):365–375
Wilkinson S, Dodgson G, Meares K (2017) Predictive processing and the varieties of psychological trauma. Frontiers in Psychology 8

Wilkinson S, Deane G, Nave K, Clark A (2019) Getting warmer: Predictive processing and the nature of emotion. In: Candiotto L (ed) The Value of Emotions for Knowledge, Springer International Publishing, Cham, pp 101–119