Burning neuroscientific questions about etiological models of psychopathy

Philip Deming

Department of Psychology, Northeastern University, Boston, MA, USA

Correspondence: p.deming@northeastern.edu

Pages: 36
Words in Abstract: 245
Words in Main Text: 4,909
Tables: 0
Figures: 1

Keywords: psychopathy; etiology; emotion; reinforcement learning; cytoarchitecture

Draft version 1.0, 9/21/2023. This paper has not been peer reviewed.

BURNING QUESTIONS ABOUT PSYCHOPATHY

Abstract

Numerous etiological models have been proposed for psychopathy, a disorder that leaves a devastating impact on society. There is a need for developing treatments that target the relevant mechanisms proposed by these models. But first, there is a need to refine the etiological models. Toward this aim, I pose three questions guided by what is known about the architecture of the human brain, and how that architecture shapes function. In question 1, I ask whether the paralimbic hypothesis implicates the whole brain. Given its myelo- and cytoarchitecture, paralimbic cortex sits atop multiple neural hierarchies. This position allows paralimbic cortex to send feedback signals down these hierarchies, shaping sensory signals from the body and world before they arrive. In question 2, I ask what the low fear theory predicts about brain function. Most previous neuroscientific applications of this theory have posited low amygdala activity as the brain basis of psychopathic individuals' purported fearlessness. I emphasize the need for a new neuroscientific application of the low fear theory, one that accounts for the distributed and context-specific brain activity that creates fear. In question 3, I ask how the brain creates goal states, and I explore implications for two models of psychopathy that assign a prominent role to goal states. Evidence from the field of reinforcement learning shows that the brain adaptively employs a spectrum of behavioral control strategies, ranging from model-free (i.e., habitual) to model-based (i.e., goal-directed). Psychopathic people may have difficulty adaptively implementing primarily model-based strategies.

Introduction

Psychopathy is a personality disorder that leaves a devastating impact on society. People with psychopathy are callous, remorseless, deceitful, and impulsive, and tend to commit repeated and varied crimes. In their wake, people with psychopathy leave victims physically and emotionally scarred and society paying a sizable bill. Kiehl & Hoffman (2011) estimated that psychopathy accounted for approximately $460 billion annually in the US in criminal social costs (e.g., costs for lost property, jails, prisons, and courts). Updated estimates put that figure as high as $1.59 trillion (Gatner et al., 2023), making psychopathy the most costly mental illness.

What causes psychopathy, and how can these causes be treated? Researchers have proposed a variety of explanations for the disorder's etiology (i.e., cause). Evidence has accumulated showing that each etiological model can explain a subset of the phenomena associated with psychopathy – for example, the low fear theory can explain psychopathic individuals' aberrant aversive startle response (Patrick, 1994) and passive avoidance learning (Lykken, 1957), but not their aberrant allocation of attention (Baskin-Sommers & Brazil, 2022). Critically, evidence-based treatments for psychopathy are lacking (although see: Baskin-Sommers et al., 2015; Olver et al., 2013), perhaps because few treatments have targeted the psychological or biological mechanisms proposed by these etiological models (Hecht et al., 2018). Thus, there is a need for (1) refining etiological models in a way that allows for a more comprehensive explanation of the broad range of psychological phenomena associated with psychopathy, and (2) empirical study of treatments designed to target theoretically relevant mechanisms.

Here I will attempt to partially address the need for refining etiological models of psychopathy by posing three questions that spring from neuroanatomy and computational neuroscience. The architecture of the brain shapes its function, and should offer clues about its dysfunction. Overall, this discussion adopts the perspective that the brain is one unified system with a complex mapping to behavior (Westlin et al., 2023). In my view, a valid and useful etiological model should be able to identify the mechanisms within this complex system that create the symptoms and behaviors of psychopathy.

Question 1: Does the paralimbic hypothesis implicate the whole brain?

The paralimbic hypothesis proposes that dysfunction in the paralimbic system (comprised of paralimbic cortex and limbic subcortical structures[1]) leads to the deficits in neurocognitive function, attention, affect, and language processing associated with psychopathy (Kiehl, 2006). This implicates a broad swath of gray matter that forms a ring within the medial aspect of the brain, including the cingulate gyrus, anterior insula, orbitofrontal cortex, parahippocampal gyrus, temporal pole, and amygdala (**Figure 1**, green areas). The gross anatomy, cytoarchitecture and myeloarchitecture of these regions make the paralimbic system uniquely positioned to have effects on activity throughout the brain. Dysfunction of the paralimbic system in psychopathy may lead to dysfunction across the whole brain.

[1] Throughout this paper I will use the same brain parcellation and nomenclature, from Mesulam (2000), that Kiehl (2006) used to propose the paralimbic hypothesis. Other parcellations and nomenclatures exist (e.g., paralimbic cortex has been called "limbic cortex"; Chanes & Barrett, 2016).

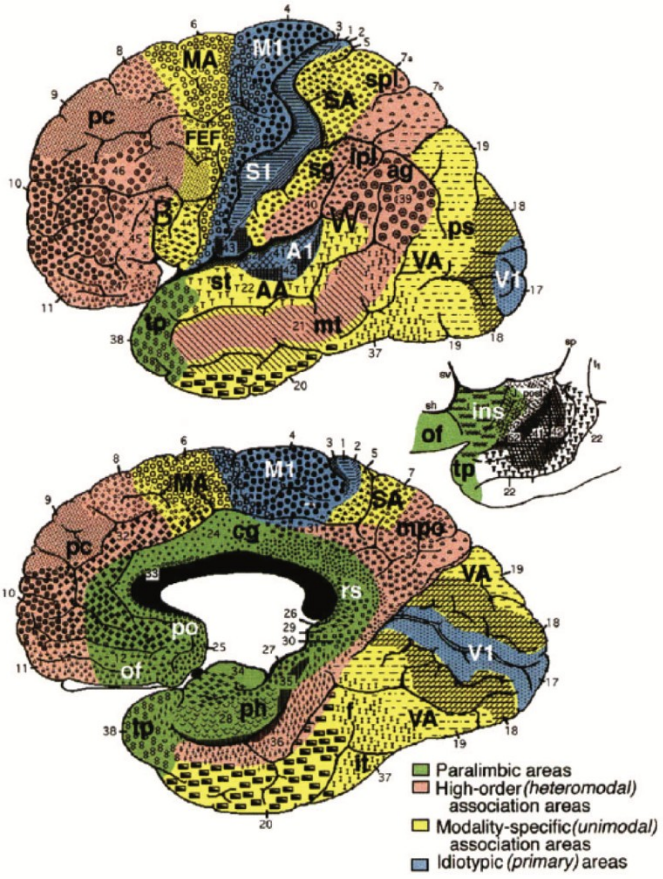

Figure 1. Functional parcellation of cortex outlined by Mesulam (2000) overlaid on Brodmann's map of the human brain. Mesulam's functional zones include paralimbic cortex (green), "heteromodal" cortex (pink), "unimodal" cortex (yellow), and "primary sensory" cortex (blue). Note that primary motor cortex (M1) is unlike the primary sensory cortices, as it lacks a layer IV (Shipp, 2005). Lateral view on top, medial view on bottom, cutout showing insula on middle right. Reproduced with permission from Mesulam (2000).

To explore this hypothesis, let us start with the architecture of the paralimbic cortex (and return to limbic subcortical structures later). The cortex is made up of 5-6 layers from superficial (layers I-III) to middle (layer IV, where thalamic projections

carrying sensory signals arrive) to deep (layers V-VI). The number and organization of these layers (i.e., the degree of lamination) changes gradually across the cortex. Paralimbic cortex is unique as it contains no layer IV (or, in some paralimbic regions, a nascent layer IV) and its layers II and III are undifferentiated (i.e., combined; Barbas, 2015; Mesulam, 2000). Moving from paralimbic cortex to what Mesulam (2000) called "heteromodal" cortex (**Figure 1**, pink areas), then to "unimodal" cortex (**Figure 1**, yellow areas) and "primary sensory" cortex (**Figure 1**, blue areas excluding M1), lamination gradually increases, as a general rule. Thus, primary sensory cortices display the highest level of lamination in the brain, containing a distinctive layer IV and differentiated layers II and III.

This gradient of lamination has implications for function. Given this gradient, paralimbic cortex has been hypothesized to sit at the top of a functional hierarchy in the cortex (Chanes & Barrett, 2016). This hypothesis is based on Barbas and colleagues' structural model of corticocortical connectivity, which has been used to accurately predict the flow of feedback (or top-down) signals and feedforward (or bottom-up) signals (Barbas, 2015; Barbas & Rempel-Clower, 1997). According to the structural model, feedback signals originate from an area with *less* lamination than the destination area, while feedforward signals originate from an area with a *greater* degree of lamination than the destination area. So the flow of feedback and feedforward signals can be traced along the gradient of cortical lamination, with feedback signals flowing from areas of less lamination (e.g., paralimbic cortex) toward areas of more lamination (e.g., primary sensory cortex), and feedforward signals flowing in the opposite direction. In this way, paralimbic cortex sends feedback signals down the hierarchy toward

primary sensory cortex. Feedback signals from paralimbic cortex likely carry multimodal contextual information (Gilbert & Li, 2013; Muckli et al., 2015; Smiley & Falchier, 2009) and modulate the firing of neurons in primary sensory cortex *before* sensory signals arrive there from peripheral sensory receptors (Barrett, 2017; Chanes & Barrett, 2016). Thus, feedback signals shape how the brain senses the world.

This functional and structural hierarchy can also be traced along white matter pathways. In broad strokes, paralimbic cortex can be thought of as 1-3 synapses from primary sensory cortex (Mesulam, 2000). The most extensive feedback projections from paralimbic cortex arrive at heteromodal cortex (as well as limbic subcortical structures; Mesulam, 2000). From heteromodal cortex, the most extensive feedback projections arrive at unimodal cortex, and from there the most extensive feedback projections arrive at primary sensory cortex. Thus, feedback signals originating from paralimbic cortex cascade down the hierarchy to arrive at primary sensory cortex. In addition, paralimbic cortex and limbic subcortical structures send monosynaptic projections to primary sensory cortex, although these projections are far less extensive than the step-wise, cascading projections down the hierarchy (Mesulam, 2000). In short, paralimbic cortex is wired to have cascading effects on activity throughout the brain.

Although subcortical structures have yet to be included in the structural model of corticocortical connectivity, evidence from tract-tracing studies suggests that limbic subcortical structures (including the amygdala and hippocampus) sit one step below paralimbic cortex in the structural hierarchy (Chanes & Barrett, 2016; Mesulam, 2000). Paralimbic cortex and limbic subcortical structures are densely interconnected (Mesulam, 2000), with limbic subcortical structures apparently sending feedforward

signals to and receiving feedback signals from paralimbic cortex (Chanes & Barrett, 2016). Feedforward signals from limbic subcortical structures are likely to carry signals from the viscera (i.e., interoceptive signals), given the dense reciprocal connections between these structures and the hypothalamus (Mesulam, 2000) as well as interoceptive brainstem structures (e.g., parabrachial nucleus; Singh et al., 2022).

To be clear, this places paralimbic cortex at the apex of two apparent hierarchies, one conveying exteroceptive (e.g., visual, auditory) sensory signals up the cortical gradient and one conveying interoceptive sensory signals up a gradient comprised of brainstem structures, hypothalamus, and limbic subcortical structures. Areas of paralimbic cortex have also been identified as “rich club” hubs with many functional and structural connections to disparate networks (van den Heuvel & Sporns, 2013). Thus, paralimbic cortex is in a prime location for integrating exteroceptive and interoceptive sensory information into multimodal summaries, and may serve as a general workspace enabling a unified experience, especially of affect (Chanes & Barrett, 2016; Menon & Uddin, 2010; Singer et al., 2009). Although, note that sensory integration is likely occurring at each step of the hierarchy (Barrett, 2017; Smith et al., 2017).

The architecture of the brain leads to this hypothesis: dysfunction in paralimbic cortex, as posited by the paralimbic hypothesis, should have effects via feedback signals on sensory processing and integration at lower stages of the neural hierarchies. Impairments in sensory processing and integration have been proposed by the impaired integration theory of psychopathy (Hamilton et al., 2015). Such impairments have been largely unexplored empirically, although one study found no evidence for impaired visual-visual integration (Gunschera et al., 2023). No study to my knowledge has

examined how people with psychopathy integrate interoceptive and exteroceptive signals, a process that may be important for creating affective experience (Chanes & Barrett, 2016; Singer et al., 2009). Future studies could draw from a variety of existing behavioral paradigms to study how psychopathic people gate exteroception via interoception (Ren et al., 2022; Salomon et al., 2016) and integrate interoceptive and exteroceptive signals (Monti et al., 2020; Suzuki et al., 2013).

In neuroimaging studies, dysfunction in the paralimbic system could have effects via feedback signals on activity in widespread regions. In fact, neuroimaging studies of psychopathy have observed task-based functional anomalies in regions scattered across all four cortical lobes and many subcortical structures (Koenigs et al., 2011). Based on the paralimbic hypothesis and neuroanatomy, one viable hypothesis is that activity of the paralimbic cortex drives these widespread anomalies. Studies might employ effective connectivity methods, such as Granger causality or dynamic causal modeling, to examine the causal effect of paralimbic cortex on activity in other regions. One initial effective connectivity study found that activity of the posterior insula (albeit a cortical region with a relatively high degree of lamination) had a reduced causal influence on amygdala activity in relation to psychopathy (Ye et al., 2022).

Limitations of this discussion include the broad-strokes, generalized portrayal of the structural and functional organization of the brain, and the exclusion of important thalamocortical connections. Nevertheless, the architecture outlined in this section can guide hypothesis development for future studies of behavior (e.g., sensory integration) and the brain (e.g., the direction of signal flow within brain networks).

Question 2: What does the low fear theory predict about brain function?

In its original form, the low fear theory made hypotheses about behavior, not brain function. The low fear theory proposes that psychopathy is caused by an inability to experience fear in anticipation of a threat or punishment (Lykken, 1957, 1995). Nonetheless, the theory has since evolved and been translated into hypotheses about brain function. Patrick (1994) first suggested a neural correlate for the proposed fearlessness, based on the contemporaneous consensus of how the brain creates[2] fear. The amygdala had long been billed as the "fear center" of the brain (Davis, 1992; LeDoux, 2020), dating back to Klüver and Bucy's observation that ablating rhesus monkeys' amygdalae led to a cessation of behaviors thought to index fear, among other changes in behavior (Klüver & Bucy, 1937; Marlowe et al., 1975). If the amygdala creates subjective experiences of fear and supports Pavlovian fear conditioning, and if people with psychopathy are unable to experience fear, then people with psychopathy must have less active amygdalae, or so the logic went. Since Patrick's initial proposal, dozens of fMRI studies have followed this logic to search for amygdala dysfunction in psychopathy (Deming et al., 2022), and several related etiological models have elaborated the hypothesis (Blair, 2003, 2005; Kiehl, 2006; Marsh, 2017; Moul et al., 2012).

However, several steps in this logic are not supported by the data. Amygdala activity is not necessary for creating a subjective experience of fear (Barrett, 2018; Feinstein et al., 2013; LeDoux, 2020), and the human amygdala may not be involved in Pavlovian fear conditioning, in contrast to findings from non-human animal studies (Fullana et al., 2016, 2018; Visser et al., 2021). Even patients with no amygdala tissue

[2] I will use the terms "create" and "cause" to ground this discussion in the primary question about how the brain creates an experience of fear, although all reviewed studies demonstrate only correlation between brain and behavior.

report experiencing fear in certain contexts, for example when inhaling elevated levels of CO_2 (Feinstein et al., 2013). Together, these findings clearly break the logic above, and suggest that a more distributed neural system creates experiences of fear.

Furthermore, the data do not support the hypothesis that psychopathy is related to chronic underactivity of the amygdala. If psychopathy were caused by chronic underactivity of the amygdala, then patients with no amygdala tissue should present with high levels of psychopathic traits. That is not the case (Lilienfeld et al., 2016). One might dismiss this finding with the following argument: the patient studied by Lilienfeld et al. (2016) suffered amygdala tissue loss around the age of 10 due to Urbach-Wiethe disease; therefore, normal fear learning may have occurred before tissue loss and distributed compensatory brain mechanisms for creating fear may have developed after tissue loss. I will present evidence below demonstrating that distributed brain mechanisms for creating fear are not compensatory, but normal.

Moreover, two decades' worth of fMRI studies of psychopathy have found inconsistent, and even contradictory, evidence for underactivity of the amygdala during experimental tasks. Stimuli thought to induce fear, including prototypical facial emotion expressions, aversive scenes, and images of another person in pain, have failed to elicit amygdala underactivity consistently across studies (Deming et al., 2022; for a meta-analysis of adolscents, see Berluti et al., 2023). Three-quarters of studies have found a null psychopathy-amygdala relationship in at least one condition. The one-third of studies that have found amygdala underactivity tended to have low study power, and most reported peak differences nearby rather than within the amygdala. One contradictory pair of findings illustrates the lack of evidence for amygdala underactivity

as a neural correlate of psychopathic individuals' purported fearlessness: one fear conditioning study administered painful pressure and found reduced amygdala activity (Birbaumer et al., 2005), while another administered electric shocks and found *increased* amygdala activity (Schultz et al., 2016). In all, the field has failed to consistently observe amygdala underactivity when psychopathic people perform tasks that are meant to induce fear.

The low fear theory is in need of a new neuroscientific application. I will not attempt to provide a full account here. But I will offer the following discussion as a way of guiding the generation of new hypotheses about the brain basis of psychopathic individuals' purported fearlessness.

Recent studies suggest the brain basis of fear is more distributed and context-specific than previously thought. Activity related to experiences of fear tends to be spread across multiple levels of the brain's hierarchies, from paralimbic regions (anterior and posterior cingulate cortex, anterior insula) down to heteromodal (dorsomedial and ventrolateral prefrontal cortex), unimodal (temporal cortex), and primary sensory regions (primary visual cortex), as well as thalamus, limbic subcortical regions (amygdala), hypothalamus, and brainstem nuclei (Kober et al., 2008; Lindquist et al., 2012; Wager et al., 2015; Wang et al., 2022). Fear is not unique in this regard: activity distributed across the whole brain likely supports all psychological phenomena, not just fear (Aliko et al., 2023; Westlin et al., 2023). Furthermore, fear appears to be created by different distributed patterns depending on the context (Wang et al., 2022). The regions involved in creating an instance of fear will likely depend on the prior state of the body (Barrett, 2017), the features of the stimulus (e.g., sensory modality), and the motor plans

generated by the perceiver (Kober et al., 2008). In each instance, subnetworks may contribute to the experience of fear by sensing and regulating the viscera (Kleckner et al., 2017), representing affect (Lindquist et al., 2016), or categorizing the instance as one of fear (Lindquist et al., 2014; Satpute & Lindquist, 2019).

Given this evidence, fear may not be a biologically prepared state with a dedicated neural circuit. Fear may be better conceptualized as a category of instances of experience that share a functional outcome (Barrett, 2017), for example escaping danger. Importantly, the perceiver, not an experimenter or an external "fear-inducing" cue, determines whether an instance is categorized as fear. No cue is inherently fear-inducing for all people. This stands in sharp contrast to classical views that fear is a universal, biologically prepared state (Ekman, 1992; Panksepp et al., 2011). Hypotheses generated from this classical view tend to garner support only under certain experimental conditions (Barrett, 2022; Gendron et al., 2020; Hoemann et al., 2023). Instead, experiences of fear are highly variable and depend on the perceiver and context. For example, how people describe the subjective experience of fear varies depending on their culture and geographic setting (Hoemann et al., 2023). There is also extensive variability in the patterns of autonomic nervous system activity related to subjective experiences of fear (Hoemann et al., 2020; Quigley & Barrett, 2014; Siegel et al., 2018). One study demonstrated that these patterns depended on the context, that is, whether the perceiver was viewing a video depicting spiders, heights, or a social situation (McVeigh et al., 2022).

If the low fear theory is to remain a viable etiological model of psychopathy, it will need to account for the context-specific relationship between brain activity and the

subjective experience of fear. Context specificity of brain function may require an adjustment to the low fear theory. If different brain mechanisms support the experience of fear in different contexts, there may be contexts in which psychopathic individuals experience fear in a way that resembles non-psychopathic people. The data suggest there are (Baskin-Sommers et al., 2011; Hoppenbrouwers et al., 2016; Newman et al., 2010). This leads to at least two new hypotheses about the brain basis of psychopathic individuals' purported fearlessness. First, there may be dysfunction in a domain-general brain mechanism that facilitates switching whole-brain activity to a state creating an instance of fear in some contexts. There is initial evidence for dysfunction in a state switching mechanism, although the data were collected during periods of rest, rather than periods marked by subjective experience of fear (Deming et al., 2023; Espinoza et al., 2019). Second, there may be dysfunction in one or multiple distributed brain mechanisms that are involved in creating fear in specific contexts. Critically, these mechanisms may differ by person and by context. Personalized analytic approaches may help to shed light on the context-specific brain basis of the fearlessness associated with psychopathy (Kraus et al., 2023; Laumann et al., 2023).

Question 3: How does the brain create goal states, according to the attention bottleneck model and motivational framework of psychopathy?

Two etiological models of psychopathy assign a prominent role to goal states. The attention bottleneck model proposes that a psychopathic person engaged in goal-directed behavior fails to attend to cues that are peripheral to their primary goal, such as the threat of punishment (Baskin-Sommers & Brazil, 2022; Wolf et al., 2012). On the other hand, the motivational framework of psychopathy proposes that psychopathic

people lack motivation to attend to stimuli that are typically considered aversive (Groat & Shane, 2020). The two etiological models are variations on a shared hypothesis: that psychopathy is marked by an altered ability to execute goal-directed behavior. Here, I will attempt to ground these models in reinforcement learning theory, a computational neuroscience approach to explaining the dynamic process by which the brain selects, implements, and navigates goal states to control behavior.

The brain has multiple systems and strategies for controlling behavior (O'Doherty, 2015; O'Doherty et al., 2017). Historically, scientists have contrasted habitual strategies, which are observable when a person selects an action based primarily on the presence of a previously reinforced stimulus (Thorndike, 1898), with goal-directed strategies, which are observable when a person acts in a way that appears to be informed by and directed toward the current expected value of an outcome (Tolman, 1948). Goal-directed strategies are thought to rely on a "cognitive map," or a representational template that enables the person to select the action expected to produce the best outcome in the moment (Tolman, 1948).

The modern field of reinforcement learning has refined these historical concepts with mathematical formalism and has stimulated discovery of the brain systems that implement different behavioral control strategies (Dolan & Dayan, 2013). In reinforcement learning language, habitual strategies are "model-free" whereas goal-directed strategies are "model-based," relying on an internal model (similar to a cognitive map) of states, potential actions, transitions between states occasioned by the potential actions, and values (Daw et al., 2005; Dolan & Dayan, 2013; O'Doherty et al., 2017; Sharpe et al., 2019). I will use reinforcement learning terminology going forward.

Importantly, model-free and model-based strategies likely make up two ends of a spectrum, rather than two discrete classes (Dolan & Dayan, 2013; Pezzulo et al., 2013). At any moment, the brain selects from this spectrum of control strategies adaptively based on current resources (O'Doherty et al., 2017).

When the brain selects model-based strategies, it seems to do so in a dynamic and distributed fashion. Although it is not fully known how the brain implements an internal model (in the reinforcement learning sense), existing evidence suggests it achieves this via dynamic interactions between multiple levels of the brain's hierarchies, from paralimbic (orbitofrontal cortex) down to heteromodal regions (posterior parietal cortex and dorsolateral prefrontal cortex), limbic subcortical structures (hippocampus), and caudate (O'Doherty et al., 2017; Sharpe et al., 2019; Wilson et al., 2014). The hippocampus has been proposed to relay information about the general structure of the environment to frontal regions (Sharpe et al., 2019), while posterior parietal cortex appears to encode the potential transitions between states depending on specific actions (O'Doherty et al., 2017). In return, dorsolateral prefrontal cortex may exert top-down attentional control to select the relevant internal model, and orbitofrontal cortex may convey a "you are here" signal representing the current state within that internal model (Sharpe et al., 2019; Wilson et al., 2014). These regions influence activity of the caudate, which plays a critical role in selecting motor actions in the current state (Sharpe et al., 2019). This process also requires a representation of the identity of the outcome, which has been attributed to orbitofrontal cortex and the basolateral amygdala (O'Doherty et al., 2017). In order to navigate the internal model (i.e., select actions with the best possible outcome in the current state), the brain represents expected values in

widespread cortical (e.g., orbitofrontal, ventromedial prefrontal, parietal, premotor, and dorsal frontal areas) and subcortical regions (e.g., amygdala and ventral and dorsal striatum; O'Doherty et al., 2017). Crucially, the brain constructs value and an internal model flexibly in the moment (O'Doherty et al., 2017). Model-based strategies enable a person to flexibly and adaptively select an action with the best expected outcome in the face of changing contingencies (i.e., possible transitions), albeit at a metabolic and computational cost.

By contrast, model-free strategies enable a person to perform previously-reinforced actions efficiently and rapidly. Though, these strategies suffer from being overly rigid and insensitive to changes in outcome values. Model-free strategies rely critically on dopaminergic brainstem nuclei and the putamen (O'Doherty et al., 2017). Model-free signals have also been observed in cortical regions such as supplementary motor area and posterior parietal cortex (Lee et al., 2014; O'Doherty et al., 2017). Within this circuit, reward prediction error signals likely serve as the mechanism for learning action-outcome associations. The brain may even adopt model-free strategies by default and switch to more computationally expensive model-based strategies only when needed (O'Doherty et al., 2017). In support of this hypothesis, ventrolateral prefrontal cortex and frontopolar cortex have been found to arbitrate between model-free and model-based systems, possibly by inhibiting the model-free system (Lee et al., 2014). These frontal regions may select a model-based strategy when that strategy has higher relative accuracy for predicting which actions should be selected (Daw et al., 2005), is not too metabolically or computationally expensive (FitzGerald et al., 2014), or substantially improves value estimation (Pezzulo et al., 2013).

According to the attention bottleneck model, when psychopathic people use model-based strategies their ability to incorporate sensory information and multiple contingencies into the internal model is limited (Baskin-Sommers & Brazil, 2022). In other words, psychopathic individuals may generate relatively sparse internal models. The theory attributes the sparseness of the internal models to excessive top-down control, mediated in part by ventrolateral prefrontal cortex (Baskin-Sommers & Brazil, 2022). Excessive top-down feedback signals may select a narrow set of sensory features of states, a narrow set of potential actions, and a narrow set of transitions between states occasioned by those actions. For example, a psychopathic person may implement a model-based strategy to obtain money, and select an internal model comprised of a narrow set of states (e.g., money in a bank vault), actions (e.g., taking), and transitions related to obtaining money. In navigating such a sparse internal model that omits states, actions, and transitions related to, for example, frightening a bank teller, the psychopathic person may select actions that amount to robbing a bank.

However, another possibility, not proposed by the attention bottleneck model, is that psychopathy is marked by an impaired capacity to shift flexibly between primarily model-free and primarily model-based strategies. In support of this hypothesis, psychopathic individuals tend to select actions that are no longer rewarded (Budhani et al., 2006; Dargis et al., 2017; Mitchell et al., 2002), which is considered a hallmark of model-free strategy use (Dolan & Dayan, 2013). This hypothesis also implicates the ventrolateral prefrontal cortex, as well as frontopolar cortex, which play a role in arbitrating between more model-free and more model-based strategies, potentially by inhibiting model-free systems (Lee et al., 2014). This hypothesis might differ from the

attention bottleneck model in terms of predicting functional brain alterations. The attention bottleneck model might predict that the ventrolateral prefrontal cortex would downregulate activity in primary sensory cortex and regions that contribute to the model-based system (e.g., hippocampus, posterior parietal cortex) when psychopathic individuals are using model-based strategies. In contrast, the inflexible shifting hypothesis might predict that ventrolateral prefrontal cortex and frontopolar cortex would fail to downregulate activity in regions that contribute to the model-free system (e.g., dopaminergic brainstem nuclei, putamen) when psychopathic individuals should be switching between model-free and model-based strategies.

The motivational framework provides a different account of psychopathic people's altered internal models. By this account, psychopathic individuals assign more neutral value to stimuli and outcomes, especially those to which non-psychopathic people tend to assign negative value (i.e., most people consider aversive; Groat & Shane, 2020). Assigning more neutral value might cause psychopathic individuals to lack motivation to attend to stimuli that most people consider aversive (e.g., another person's cry of distress). Interestingly, the motivational framework also proposes that psychopathic individuals might assign more *negative* value to such stimuli, leading to a motivation to avoid. Unlike the attention bottleneck model, which proposes that psychopathic individuals generate sparse internal models, the motivational framework posits that psychopathic individuals using model-based strategies are navigating an internal model that simply contains value assignments that largely differ from the value assignments that most people tend to learn. Specifically, these value assignments are thought to lead them to avoid states of negative valence (although see Spantidaki

Kyriazi et al., 2020). Instead, their unique value assignments might lead them toward outcomes that confer pleasure, wealth, or a higher position in society (Glenn et al., 2017). The motivational framework does not offer a mechanistic account of value learning in psychopathy, although other researchers have weighed in (Blair, 2013; Moul et al., 2012). I would attribute their altered value learning to impaired integration of interoceptive and exteroceptive signals, given that interoceptive signals may be integral to learning the value of sensory states and actions (Damasio et al., 1996; Gu & FitzGerald, 2014). This hypothesis remains to be studied.

In sum, human brains flexibly and adaptively employ a spectrum of behavioral control strategies, ranging from model-free to model-based. People with psychopathy may have unique impairments in implementing model-based strategies, potentially characterized by excessive top-down control resulting in a sparse internal model (Baskin-Sommers & Brazil, 2022), inflexible shifting between model-free and model-based strategies, altered value learning (Blair, 2013; Groat & Shane, 2020; Moul et al., 2012), or some combination thereof. More empirical work embedded in the reinforcement learning framework (e.g., Driessen et al., 2021; Oba et al., 2019) is needed to characterize how psychopathic individuals construct internal models, and how and when they select model-based strategies to control behavior.

General Conclusions

The practice of checking an etiological model against neuroanatomy and computational neuroscience helps to refine hypotheses not only about brain dysfunction, but also about cognition and behavior. In particular, scientists proposing etiological models would do well to adopt the perspective that the human brain is one

complex system that flexibly switches between states (McCormick et al., 2020) and strategies for controlling behavior (O'Doherty et al., 2017). A hierarchical structure from multimodal paralimbic cortex down to primary sensory cortex all the way down to sensory receptors (Chanes & Barrett, 2016; Mesulam, 2000) forms the contours along which the brain makes emotional meaning (via feedback signals) out of sensory information (Barrett, 2017) and selects strategies for controlling behavior. Moreover, the mapping between brain and behavior differs across contexts (Westlin et al., 2023). Consequently, the mechanisms and deficits associated with psychopathy may differ across people and across contexts depending, for example, on the state of brain activity in the moment prior (McCormick et al., 2020), the brain's estimation of the body's metabolic state (FitzGerald et al., 2014), or the complexity of sensory states and contingencies at the moment (Baskin-Sommers & Brazil, 2022). Therefore, this perspective has the potential to advance our understanding of heterogeneous symptom presentation across psychopathic people, and also to integrate across etiological models in a way that eschews distinctions between psychological phenomena such as attention and emotion. Further research is needed to characterize how the structural and functional hierarchy of a psychopathic person's brain instantiates context-specific experiences of fear and internal models for guiding model-based control strategies.

Acknowledgments

I would like to thank Alexandra Fischbach and Kieran McVeigh for providing constructive feedback on an initial draft of this paper.

References

Aliko, S., Wang, B., Small, S. L., & Skipper, J. I. (2023). *The entire brain, more or less is at work: "Language regions" are artefacts of averaging* (p. 2023.09.01.555886). bioRxiv. https://doi.org/10.1101/2023.09.01.555886

Barbas, H. (2015). General Cortical and Special Prefrontal Connections: Principles from Structure to Function. *Annual Review of Neuroscience*, *38*(1), 269–289. https://doi.org/10.1146/annurev-neuro-071714-033936

Barbas, H., & Rempel-Clower, N. (1997). Cortical structure predicts the pattern of corticocortical connections. *Cerebral Cortex*, *7*(7), 635–646. https://doi.org/10.1093/cercor/7.7.635

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsw154

Barrett, L. F. (2018). Seeing Fear: It's All in the Eyes? *Trends in Neurosciences*, *41*(9), 559–563. https://doi.org/10.1016/j.tins.2018.06.009

Barrett, L. F. (2022). Context Reconsidered: Complex Signal Ensembles, Relational Meaning, and Population Thinking in Psychological Science. *American Psychologist*, *77*(8), 894–920.

Baskin-Sommers, A. R., & Brazil, I. A. (2022). The importance of an exaggerated attention bottleneck for understanding psychopathy. *Trends in Cognitive Sciences*, *26*(4), 1–12. https://doi.org/10.1016/j.tics.2022.01.001

Baskin-Sommers, A. R., Curtin, J. J., & Newman, J. P. (2011). Specifying the attentional selection that moderates the fearlessness of psychopathic offenders.

Psychological Science, *22*(2), 226–234. https://doi.org/10.1177/0956797610396227

Baskin-Sommers, A. R., Curtin, J. J., & Newman, J. P. (2015). Altering the Cognitive-Affective Dysfunctions of Psychopathic and Externalizing Offender Subtypes With Cognitive Remediation. *Clinical Psychological Science*, *3*(1), 45–57. https://doi.org/10.1177/2167702614560744

Berluti, K., Ploe, M. L., & Marsh, A. A. (2023). Emotion processing in youths with conduct problems: An fMRI meta-analysis. *Translational Psychiatry*, *13*(1), 105. https://doi.org/10.1038/s41398-023-02363-z

Birbaumer, N., Veit, R., Lotze, M., Erb, M., Hermann, C., Grodd, W., & Flor, H. (2005). Deficient Fear Conditioning in Psychopathy: A Functional Magnetic Resonance Imaging Study. *Archives of General Psychiatry*, *62*(7), 799–805. https://doi.org/10.1001/archpsyc.62.7.799

Blair, R. J. R. (2003). Neurobiological basis of psychopathy. *The British Journal of Psychiatry*, *182*(1), 5–7. https://doi.org/10.1192/bjp.182.1.5

Blair, R. J. R. (2005). Applying a cognitive neuroscience perspective to the disorder of psychopathy. *Development and Psychopathology*, *17*, 865–891.

Blair, R. J. R. (2013). Psychopathy: Cognitive and neural dysfunction. *Dialogues in Clinical Neuroscience*, *15*(2), 181–190. https://doi.org/10.31887/DCNS.2013.15.2/rblair

Budhani, S., Richell, R. A., & Blair, R. J. R. (2006). Impaired reversal but intact acquisition: Probabilistic response reversal deficits in adult individuals with

psychopathy. *Journal of Abnormal Psychology*, *115*(3), 552–558. https://doi.org/10.1037/0021-843X.115.3.552

Chanes, L., & Barrett, L. F. (2016). Redefining the Role of Limbic Areas in Cortical Processing. *Trends in Cognitive Sciences*, *20*(2), 96–106. https://doi.org/10.1016/j.tics.2015.11.005

Damasio, A. R., Everitt, B. J., & Bishop, D. (1996). The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex [and Discussion]. *Philosophical Transactions: Biological Sciences*, *351*(1346), 1413–1420.

Dargis, M., Wolf, R. C., & Koenigs, M. R. (2017). Reversal Learning Deficits in Criminal Offenders: Effects of Psychopathy, Substance use, and Childhood Maltreatment History. *Journal of Psychopathology and Behavioral Assessment*, *39*(2), 189–197.

Davis, M. (1992). The Role of the Amygdala in Fear and Anxiety. *Annual Review of Neuroscience*, *15*(1), 353–375. https://doi.org/10.1146/annurev.ne.15.030192.002033

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), Article 12. https://doi.org/10.1038/nn1560

Deming, P., Cook, C. J., Meyerand, M. E., Kiehl, K. A., Kosson, D. S., & Koenigs, M. (2023). Impaired salience network switching in psychopathy. *Behavioural Brain Research*, *452*, 114570. https://doi.org/10.1016/j.bbr.2023.114570

Deming, P., Heilicher, M., & Koenigs, M. (2022). How reliable are amygdala findings in psychopathy? A systematic review of MRI studies. *Neuroscience & Biobehavioral Reviews*, 104875. https://doi.org/10.1016/j.neubiorev.2022.104875

Dolan, R. J., & Dayan, P. (2013). Goals and Habits in the Brain. *Neuron*, *80*(2), 312–325. https://doi.org/10.1016/j.neuron.2013.09.007

Driessen, J. M. A., van Baar, J. M., Sanfey, A. G., Glennon, J. C., & Brazil, I. A. (2021). Moral strategies and psychopathic traits. *Journal of Abnormal Psychology*, *130*(5), 550–561. https://doi.org/10.1037/abn0000675

Ekman, P. (1992). Are there basic emotions? *Psychological Review*, *99*(3), 550–553.

Espinoza, F. A., Anderson, N. E., Vergara, V. M., Harenski, C. L., Decety, J., Rachakonda, S., Damaraju, E., Koenigs, M. R., Kosson, D. S., Harenski, K. A., Calhoun, V. D., & Kiehl, K. A. (2019). Resting-state fMRI dynamic functional network connectivity and associations with psychopathy traits. *NeuroImage: Clinical*, *24*(July), 101970. https://doi.org/10.1016/j.nicl.2019.101970

Feinstein, J. S., Buzza, C., Hurlemann, R., Follmer, R. L., Dahdaleh, N. S., Coryell, W. H., Welsh, M. J., Tranel, D., & Wemmie, J. A. (2013). Fear and panic in humans with bilateral amygdala damage. *Nature Neuroscience*, *16*(3), 270–272. https://doi.org/10.1038/nn.3323

FitzGerald, T. H. B., Dolan, R. J., & Friston, K. J. (2014). Model averaging, optimal inference, and habit formation. *Frontiers in Human Neuroscience*, *8*. https://www.frontiersin.org/articles/10.3389/fnhum.2014.00457

Fullana, M. A., Albajes-Eizagirre, A., Soriano-Mas, C., Vervliet, B., Cardoner, N., Benet, O., Radua, J., & Harrison, B. J. (2018). Fear extinction in the human brain: A

meta-analysis of fMRI studies in healthy participants. *Neuroscience and Biobehavioral Reviews*, *88*(February), 16–25. https://doi.org/10.1016/j.neubiorev.2018.03.002

Fullana, M. A., Harrison, B. J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., & Radua, J. (2016). Neural signatures of human fear conditioning: An updated and extended meta-analysis of fMRI studies. *Molecular Psychiatry*, *21*(4), 500–508. https://doi.org/10.1038/mp.2015.88

Gatner, D. T., Douglas, K. S., Almond, M. F. E., Hart, S. D., & Kropp, P. R. (2023). How much does that cost? Examining the economic costs of crime in North America attributable to people with psychopathic personality disorder. *Personality Disorders: Theory, Research, and Treatment*, *14*(4), 391–400. https://doi.org/10.1037/per0000575

Gendron, M., Hoemann, K., Crittenden, A. N., Mangola, S. M., Ruark, G. A., & Barrett, L. F. (2020). Emotion Perception in Hadza Hunter-Gatherers. *Scientific Reports*, 1–17. https://doi.org/10.1038/s41598-020-60257-2

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350–363. https://doi.org/10.1038/nrn3476

Glenn, A. L., Efferson, L. M., Iyer, R., & Graham, J. (2017). Values, Goals, and Motivations Associated with Psychopathy. *Journal of Social and Clinical Psychology*, *36*(2), 108–125. https://doi.org/10.1521/jscp.2017.36.2.108

Groat, L. L., & Shane, M. S. (2020). A Motivational Framework for Psychopathy: Toward a Reconceptualization of the Disorder. *European Psychologist*, *25*(2), 92–103. https://doi.org/10.1027/1016-9040/a000394

Gu, X., & FitzGerald, T. H. B. (2014). Interoceptive inference: Homeostasis and decision-making. *Trends in Cognitive Sciences*, *18*(6), 269–270. https://doi.org/10.1016/j.tics.2014.02.001

Gunschera, L. J., Verschuere, B., Murphy, R. A., Temple-McCune, A., Dutton, K., & Fox, E. (2023). No impaired integration in psychopathy: Evidence from an illusory conjunction paradigm. *Personality Disorders: Theory, Research, and Treatment*. https://doi.org/10.1037/per0000619

Hamilton, R. K. B., Racer, K. H., & Newman, J. P. (2015). Impaired Integration in Psychopathy: A Unified Theory of Psychopathic Dysfunction. *Psychological Review*, *122*(4), 770–791.

Hecht, L. K., Latzman, R. D., & Lilienfeld, S. O. (2018). The Psychological Treatment of Psychopathy: Theory and Research. In D. David, S. J. Lynn, & G. H. Montgomery (Eds.), *Evidence-Based Psychotherapy: The State of the Science and Practice* (pp. 271–298). Wiley-Blackwell. https://doi.org/10.1002/9781119462996.ch11

Hoemann, K., Gendron, M., Crittenden, A. N., Mangola, S. M., Endeko, E. S., Dussault, È., Barrett, L. F., & Mesquita, B. (2023). What We Can Learn About Emotion by Talking With the Hadza. *Perspectives on Psychological Science*, 17456916231178555. https://doi.org/10.1177/17456916231178555

Hoemann, K., Khan, Z., Feldman, M. J., Nielson, C., Devlin, M., Dy, J., Barrett, L. F., Wormwood, J. B., & Quigley, K. S. (2020). Context-aware experience sampling reveals the scale of variation in affective experience. *Scientific Reports*, *10*(1), 1–17. https://doi.org/10.1038/s41598-020-69180-y

Hoppenbrouwers, S. S., Bulten, B. H., & Brazil, I. A. (2016). Parsing Fear: A Reassessment of the Evidence for Fear Deficits in Psychopathy. *Psychological Bulletin*, *142*(6), 1–29. https://doi.org/10.1037/bul0000040

Kiehl, K. A. (2006). A cognitive neuroscience perspective on psychopathy: Evidence for paralimbic system dysfunction. *Psychiatry Research*, *142*, 107–128. https://doi.org/10.1016/j.psychres.2005.09.013.A

Kiehl, K. A., & Hoffman, M. B. (2011). The Criminal Psychopath: History, neuroscience, treatment, and economics. *Jurimetrics*, *51*, 355–397. http://dx.doi.org/10.1108/17506200710779521

Kleckner, I. R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W. K., Quigley, K. S., Dickerson, B. C., & Feldman Barrett, L. (2017). Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nature Human Behaviour*, *1*, 1–14. https://doi.org/10.1038/s41562-017-0069

Klüver, H., & Bucy, P. C. (1937). " Psychic blindness" and other symptoms following bilateral temporal lobectomy in Rhesus monkeys. *American Journal of Physiology*.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K. A., & Wager, T. D. (2008). Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *NeuroImage*, *42*, 998–1031. https://doi.org/10.1016/j.neuroimage.2008.03.059

Koenigs, M. R., Baskin-Sommers, A. R., Zeier, J., & Newman, J. P. (2011). Investigating the neural correlates of psychopathy: A critical review. *Molecular Psychiatry*, *16124*, 792–799. https://doi.org/10.1038/mp.2010.124

Kraus, B., Zinbarg, R., Braga, R. M., Nusslock, R., Mittal, V. A., & Gratton, C. (2023). Insights from personalized models of brain and behavior for identifying biomarkers in psychiatry. *Neuroscience & Biobehavioral Reviews*, *152*, 105259. https://doi.org/10.1016/j.neubiorev.2023.105259

Laumann, T. O., Zorumski, C. F., & Dosenbach, N. U. F. (2023). Precision Neuroimaging for Localization-Related Psychiatry. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2023.1576

LeDoux, J. E. (2020). Thoughtful feelings. *Current Biology*, *30*(11), R619–R623. https://doi.org/10.1016/j.cub.2020.04.012

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. *Neuron*, *81*(3), 687–699. https://doi.org/10.1016/j.neuron.2013.11.028

Lilienfeld, S. O., Sauvigné, K. C., Reber, J., Watts, A. L., Hamann, S., Smith, S. F., Patrick, C. J., Bowes, S. M., & Tranel, D. (2016). Potential Effects of Severe Bilateral Amygdala Damage on Psychopathic Personality Features: A Case Report. *Personality Disorders: Theory, Research, and Treatment*, *9*(2), 1–10.

Lindquist, K. A., Gendron, M., Barrett, L. F., & Dickerson, B. C. (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion*, *14*(2), 375–387. https://doi.org/10.1037/a0035293

Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2016). The Brain Basis of Positive and Negative Affect: Evidence from a Meta-Analysis of the Human Neuroimaging Literature. *Cerebral Cortex*, *26*(5), 1910–1922. https://doi.org/10.1093/cercor/bhv001

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences*, *35*, 121–143. https://doi.org/10.1017/S0140525X11000446

Lykken, D. T. (1957). A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology*, *55*(1), 6–10.

Lykken, D. T. (1995). *The Antisocial Personalities*. Psychology Press.

Marlowe, W. B., Mancall, E. L., & Thomas, J. J. (1975). Complete Klüver-Bucy Syndrome in Man. *Cortex*, *11*(1), 53–59. https://doi.org/10.1016/S0010-9452(75)80020-7

Marsh, A. (2017). *The fear factor: How one emotion connects altruists, psychopaths, and everyone in-between*. Hachette UK.

McCormick, D. A., Nestvogel, D. B., & He, B. J. (2020). Neuromodulation of Brain State and Behavior. *Annual Review of Neuroscience*, *43*(1), 391–415. https://doi.org/10.1146/annurev-neuro-100219-105424

McVeigh, K., Kleckner, I. R., Quigley, K. S., & Satpute, A. B. (2022). *Fear-related psychophysiological patterns are situation and individual dependent: A Bayesian model comparison approach.* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/7uk4z

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure and Function*, 1–13. https://doi.org/10.1007/s00429-010-0262-0

Mesulam, M.-M. (2000). *Principles of behavioral and cognitive neurology*. Oxford University Press.

Mitchell, D. G. V., Colledge, E., Leonard, A., & Blair, R. J. R. (2002). Risky decisions and response reversal: Is there evidence of orbitofrontal cortex dysfunction in psychopathic individuals? In *Neuropsychologia* (Vol. 40, pp. 2013–2022).

Monti, A., Porciello, G., Tieri, G., & Aglioti, S. M. (2020). The “embreathment” illusion highlights the role of breathing in corporeal awareness. *Journal of Neurophysiology*, *123*(1), 420–427. https://doi.org/10.1152/jn.00617.2019

Moul, C., Killcross, S., & Dadds, M. R. (2012). A model of differential amygdala activation in psychopathy. *Psychological Review*, *119*(4), 789–806. https://doi.org/10.1037/a0029342

Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., Goebel, R., & Yacoub, E. (2015). Contextual Feedback to Superficial Layers of V1. *Current Biology*, *25*(20), 2690–2695. https://doi.org/10.1016/j.cub.2015.08.057

Newman, J. P., Curtin, J. J., Bertsch, J. D., & Baskin-Sommers, A. R. (2010). Attention moderates the fearlessness of psychopathic offenders. *Biological Psychiatry*, *67*(1), 66–70. https://doi.org/10.1016/j.biopsych.2009.07.035.Attention

Oba, T., Katahira, K., & Ohira, H. (2019). The Effect of Reduced Learning Ability on Avoidance in Psychopathy: A Computational Approach. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02432

O’Doherty, J. P. (2015). Multiple systems for the motivational control of behavior and associated neural substrates in humans. In E. H. Simpson & P. D. Balsam (Eds.), *Behavioral Neuroscience of Motivation* (Vol. 27, pp. 291–312). Springer.

O'Doherty, J. P., Cockburn, J., & Pauli, W. M. (2017). Learning, Reward, and Decision Making. *Annual Review of Psychology*, *68*(1), 73–100. https://doi.org/10.1146/annurev-psych-010416-044216

Olver, M. E., Lewis, K., & Wong, S. C. P. (2013). Risk reduction treatment of high-risk psychopathic offenders: The relationship of psychopathy and treatment change to violent recidivism. *Personality Disorders: Theory, Research, and Treatment*, *4*(2), 160–167. https://doi.org/10.1037/a0029769

Panksepp, J., Fuchs, T., & Iacobucci, P. (2011). The basic neuroscience of emotional experiences in mammals: The case of subcortical FEAR circuitry and implications for clinical anxiety. *Applied Animal Behaviour Science*, *129*(1), 1–17. https://doi.org/10.1016/j.applanim.2010.09.014

Patrick, C. J. (1994). Emotion and psychopathy: Startling new insights. *Psychophysiology*, *31*(4), 319–330. https://doi.org/10.1111/j.1469-8986.1994.tb02440.x

Pezzulo, G., Rigoli, F., & Chersi, F. (2013). The Mixed Instrumental Controller: Using Value of Information to Combine Habitual Choice and Mental Simulation. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00092

Quigley, K. S., & Barrett, L. F. (2014). Is there consistency and specificity of autonomic changes during emotional episodes? Guidance from the Conceptual Act Theory and psychophysiology. *Biological Psychology*, *98*, 82–94. https://doi.org/10.1016/j.biopsycho.2013.12.013

Ren, Q., Marshall, A. C., Kaiser, J., & Schütz-Bosbach, S. (2022). Multisensory integration of anticipated cardiac signals with visual targets affects their detection among multiple visual stimuli. *NeuroImage*, *262*, 119549. https://doi.org/10.1016/j.neuroimage.2022.119549

Salomon, R., Ronchi, R., Dönz, J., Bello-Ruiz, J., Herbelin, B., Martet, R., Faivre, N., Schaller, K., & Blanke, O. (2016). The Insula Mediates Access to Awareness of Visual Stimuli Presented Synchronously to the Heartbeat. *The Journal of Neuroscience*, *36*(18), 5115–5127. https://doi.org/10.1523/JNEUROSCI.4262-15.2016

Satpute, A. B., & Lindquist, K. A. (2019). The Default Mode Network's Role in Discrete Emotion. *Trends in Cognitive Sciences*, *23*(10), 851–864. https://doi.org/10.1016/j.tics.2019.07.003

Schultz, D. H., Balderston, N. L., Baskin-Sommers, A. R., Larson, C. L., & Helmstetter, F. J. (2016). Psychopaths show enhanced amygdala activation during fear conditioning. *Frontiers in Psychology*, *7*(MAR), 1–12. https://doi.org/10.3389/fpsyg.2016.00348

Sharpe, M. J., Stalnaker, T., Schuck, N. W., Killcross, S., Schoenbaum, G., & Niv, Y. (2019). An Integrated Model of Action Selection: Distinct Modes of Cortical Control of Striatal Decision Making. *Annual Review of Psychology*, *70*(1), 53–76. https://doi.org/10.1146/annurev-psych-010418-102824

Shipp, S. (2005). The importance of being agranular: A comparative account of visual and motor cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 797–814. https://doi.org/10.1098/rstb.2005.1630

Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., Quigley, K. S., & Feldman Barrett, L. (2018). Emotion Fingerprints or Emotion Populations? A Meta-Analytic Investigation of Autonomic Features of Emotion Categories. *Psychological Bulletin*, *144*(4), 343–393. https://doi.org/10.1037/bul0000128

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences*, *13*(8), 334–340. https://doi.org/10.1016/j.tics.2009.05.001

Singh, K., García-Gomar, M. G., Cauzzo, S., Staab, J. P., Indovina, I., & Bianciardi, M. (2022). Structural connectivity of autonomic, pain, limbic, and sensory brainstem nuclei in living humans based on 7 Tesla and 3 Tesla MRI. *Human Brain Mapping*, *43*(10), 3086–3112. https://doi.org/10.1002/hbm.25836

Smiley, J. F., & Falchier, A. (2009). Multisensory connections of monkey auditory cerebral cortex. *Hearing Research*, *258*(1), 37–46. https://doi.org/10.1016/j.heares.2009.06.019

Smith, R., Thayer, J. F., Khalsa, S. S., & Lane, R. D. (2017). The hierarchical basis of neurovisceral integration. *Neuroscience & Biobehavioral Reviews*, *75*, 274–296. https://doi.org/10.1016/j.neubiorev.2017.02.003

Spantidaki Kyriazi, F., Bogaerts, S., Tamir, M., Denissen, J. J. A., & Garofalo, C. (2020). Emotion Goals: A Missing Piece in Research on Psychopathy and Emotion Regulation. *Journal of Personality Disorders*, 1-S7. https://doi.org/10.1521/pedi_2020_34_488

Suzuki, K., Garfinkel, S. N., Critchley, H. D., & Seth, A. K. (2013). Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. *Neuropsychologia*, *51*(13), 2909–2917. https://doi.org/10.1016/j.neuropsychologia.2013.08.014

Thorndike, E. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review Monograph Supplements*, *2*(4), 1–109.

Tolman, E. (1948). Cognitive maps in rats and men. *Psychological Revew*, *55*(4), 189–208.

van den Heuvel, M. P., & Sporns, O. (2013). Network hubs in the human brain. *Trends in Cognitive Sciences*, *17*(12), 683–696. https://doi.org/10.1016/j.tics.2013.09.012

Visser, R. M., Bathelt, J., Scholte, H. S., & Kindt, M. (2021). Robust BOLD responses to faces but not to conditioned threat: Challenging the amygdala's reputation in human fear and extinction learning Robust BOLD responses to faces but not to conditioned threat: Challenging the amygdala ' s reputation in human fe. *Journal of Neuroscience*, *41*(50), 10278–10292.

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., & Barrett, L. F. (2015). A Bayesian Model of Category-Specific Emotional Brain Responses. *PLoS Computational Biology*, *11*(4), 1–27. https://doi.org/10.1371/journal.pcbi.1004066

Wang, Y., Kragel, P. A., & Satpute, A. B. (2022). *Neural predictors of subjective fear depend on the situation* [Preprint]. bioRxiv. https://doi.org/10.1101/2022.10.20.513114

Westlin, C., Theriault, J. E., Katsumi, Y., Nieto-Castanon, A., Kucyi, A., Ruf, S. F., Brown, S. M., Pavel, M., Erdogmus, D., Brooks, D. H., Quigley, K. S., Whitfield-Gabrieli, S., & Barrett, L. F. (2023). Improving the study of brain-behavior relationships by revisiting basic assumptions. *Trends in Cognitive Sciences*, *27*(3), 246–257. https://doi.org/10.1016/j.tics.2022.12.015

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron*, *81*(2), 267–279. https://doi.org/10.1016/j.neuron.2013.11.005

Wolf, R. C., Carpenter, R. W., Warren, C. M., Zeier, J. D., Baskin-Sommers, A. R., & Newman, J. P. (2012). Reduced susceptibility to the attentional blink in psychopathic offenders: Implications for the attention bottleneck hypothesis. *Neuropsychology*, *26*(1), 102–109. https://doi.org/10.1037/a0026000

Ye, S., Li, W., Zhu, B., Lv, Y., Yang, Q., & Krueger, F. (2022). Altered effective connectivity from the posterior insula to the amygdala mediates the relationship between psychopathic traits and endorsement of the Harm foundation. *Neuropsychologia*, *170*(March), 108216. https://doi.org/10.1016/j.neuropsychologia.2022.108216