Specific Phobia is an Ideal Psychiatric Kind

Alexander Pereira [*]

Winner of the 2020 Karl Jaspers Award[†]
Final manuscript draft (please do not cite)[‡]

Abstract

This paper argues that specific phobia is an ideal kind of psychiatric disorder because it bears the marks of a mature medical diagnosis and is amenable to causal explanation. A new and ambitious program of 'causal revolution' has recently emerged in psychiatry that hopes to refurnish our taxonomies by discovering the underlying biological and psychological causes that create and maintain mental illness. I show that the sort of causal story envisioned by the program is a mechanistic property cluster (MPC) structure, which involves a causal mechanism that explains the co-occurrence of a disorder's signs and symptoms (Kendler, Zachar & Craver, 2011). I then build a model of fear in humans and sketch a novel account of specific phobia as a configuration of the fear system in thrall to deregulated network dynamics such as hysteresis, tipping points, and feedback loops. Specific phobia has an MPC structure. I close by reflecting on whether we can reasonably expect other mental disorders to fit an MPC mold, and thus lend themselves to future causal validation. This paper shows that specific phobia holds a unique place in our picture of mental disorder that has so far been missed. It is an ideal kind of psychopathology.

KEYWORDS: Fear, mechanistic property clusters, explanation, anxiety disorders, dysregulation, networks.

1. Classificatory Pessimism and the New Causal Revolution in Psychiatry

The causes and underlying natures of common mental disorders are, for the most part, quite mysterious. Our best taxonomies acknowledge this poverty of causal knowledge about minds, brains, society, and whatever else, to instead classify psychopathology based on clusters of detectable signs and symptoms: what it is to be, say, depressed, is simply to exhibit the minimum number of typical features for the right amount of time. Nothing in this approach references what *causes* and *maintains* a characteristic set of symptoms, and so by conceptualizing disorders as syndromes psychiatry stands in stark contrast to somatic medicine, its close (but more mature) cousin, and other successful sciences that classify aetiologically.

[*] University of Sydney, School for the History and Philosophy of Science, Sydney, NSW, Australia, 2006. alexander.pereira@sydney.edu.au.

[†] Awarded by the Association for the Advancement of Philosophy of Psychiatry (AAPP). https://philosophyandpsychiatry.org/jaspers-award/.

This is somewhat forgivable. The innerworkings of psychiatric disorders[4] appear to depend on complicated and perhaps inscrutable webs of unknown interactions that span levels of organization, from intricate molecular genetics to enigmatic sociocultural forces. So, the *Diagnostic and Statistical Manual of Mental Disorder* (5th ed.; *DSM-5*; American Psychiatric Association; APA). – the "bible" of psychiatry – privileges clinical manifestation while staying agnostic about underlying cause. In this, it performs an important pragmatic role in fostering interdisciplinary communication and connecting urgently suffering people to treatment.

Worries over this classification system, however, abound. Philosophers and clinicians have criticized the DSM for being infected with folk-psychological concepts that mislead and obstruct scientific progress (Murphy, 2006); for being built within a culture of internecine academic warfare by work-groups who are especially vulnerable to social and political interference (Greenberg, 2013); and for pathologizing normal behavior (Horwitz & Wakefield, 2007). Furthermore, mental disorders as currently described are highly comorbid and prone to large individual variations (APA, 2013). This suggests that diagnostic boundaries are ill-defined, categories are confusing and conflating distinct psychiatric processes, and some disorders cause others in unknown ways. So, what looks like a true comorbidity involving the genuine co-existence of independent disorders might instead be a spurious artifact of immature classifications. Finally, many psychiatric disorders appear to manifest differently across sociocultural context. This is an especially difficult puzzle for psychiatry and it is an open question as to whether the same underlying problems are being modulated by different environmental variables, or whether society and culture play far more fundamental roles in constructing mental illness (Kleinman, 1987; cf. Horwitz & Wakefield, 2007, pp. 195-202). In this, a chasm of difference exists between the limited ability of psychiatry to furnish explanations, predictions, and interventions, and the robust ability of somatic medicine to do the same for biological disease.

In response, philosophers and practitioners have called for a revised taxonomy that discriminates disorders by causes. Indeed, in 2010 the National Institute of Mental Health (NIMH) launched the "Research Domain Criteria project" to direct novel investigations into the biological and psychological mechanisms of mental disorder that explicitly ignores the DSM, with many hoping it will act as a springboard for new, scientifically valid depictions of psychopathology. Thomas Insel, former director of the NIMH, wrote:

> "Mental disorders are biological disorders involving brain circuits that implicate specific domains of cognition, emotion and behaviour…mapping the cognitive, circuit, and genetic aspects of mental disorders will yield new and better targets for treatment…it is critical to realize we cannot succeed if we use DSM categories as the 'gold standard.'" (Insel, 2013).

[4] I use the terms psychiatric disorder, mental disorder, and mental illness interchangeably to represent our general ideas about departures from mental health. Psychopathology is reserved for the more scientifically rigorous notion of mental disorders corresponding to underlying pathogenic processes inside the psychiatric domain.

Given the slew of criticisms leveled at the DSM-5, it seems to me crucial that we focus on sorting out which of our current categories of psychiatric disorder are trustworthy and which require radical revision. This "causal revolution" asks us to attune psychiatry to whatever underlying states-of-affairs are actually performing the loadbearing disorder-generating work in creating mental illness. This is a daunting task. What interests me in this paper is what psychiatric kinds must look like to survive this causal scrutiny, and whether there are any suitable candidates lurking within our current taxonomies. This is similar to what Murphy (2014, p. 119) dubs the *vindication project* regarding the vindication of folk-psychological concepts through the discovery of an underlying "causal signature". We can recast the project here as the search for a validating causal story that promotes a DSM-defined disorder to a scientifically defensible kind.

In this paper, I argue for two points that dovetail. First, psychiatric disorders are likely mechanistic property cluster (MPC) kinds, which are clusters of properties generated and held together by a causal mechanism. Articulating the MPC structure of a putative kind of mental illness is the same as discovering its causal signature, so this structure is the target of the vindication project. Second, specific phobia is an ideal MPC kind and is thus an ideal kind of psychopathology. It is a "gold standard" DSM category.

Specific phobia involves an intense, inappropriate fear of a particular object or situation that causes significant distress and avoidance (APA, 2013). It has been largely ignored by the philosophy of psychiatry and is often downplayed to spotlight the general class "anxiety disorders." Specific phobia deserves special attention.

I show that specific phobia has unique characteristics that have been either unappreciated or completely missed by the field. Together, these give us *prima facie* reason to think the DSM depiction of specific phobia has struck a genuine MPC kind. First, phobia has a strong pancultural character suggesting it involves a common underlying process that is only minimally shaped by society and culture. Second, fear is evolutionarily ancient, reasonably circumscribed, and scientifically well-understood; if we want to account for fear dysfunction, this bedrock of cohesive (if incomplete) research sets us up nicely. Third, and most strikingly, phobia is successfully treated and often cured by a single one-size-fits-all intervention: exposure therapy. This is extraordinarily rare in psychiatry. Specific phobia is uncontroversially a *mental* disorder that nevertheless bears the marks of a mature medical disease.

So, this paper will introduce and sketch the contours of specific phobia as an MPC kind. In doing so, I argue it is an ideal psychiatric kind and precisely the sort of thing that the vindication project envisions. Here's the game plan. In Section 2, I discuss the question of what kind of things psychiatric disorders are and argue that the best model is the MPC view synthesized with a vocabulary of network dynamics. Section 3 constructs a model of the fear system in humans. Section 4 provides an MPC account of specific phobia. I argue that specific phobia arises from a broad pattern of fear dysregulation that locks the plastic fear system into inflexible cycles of intense response and powerful avoidance. In short, phobia is a configuration of the fear system in thrall to dysregulated network dynamics such as hysteresis, tipping points, and feedback loops. I close in Section 5 with a discussion of the vindication project and how reasonable it is to expect other disorders to fit the MPC mold.

2. What Kinds of Things are Mental Disorders?

2.1 Essentialism and Complexity

By asking us to attune our picture of mental disorders to the causal structure of the world, this program of psychiatric reform is simply asking us to aim our classifications at whatever is actually going on in mental disorders; to whatever mental disorders *really* are. So, what *kinds* of things are mental disorders, anyway? What sort of "causal signature" are we to look for?

One unhelpful answer is that psychiatric kinds depend on *essences* – members of a kind, like cases of a mental illness, are united because they share an "essential" internal component that creates the distinguishing features of the kind. Gold's unique atomic structure (79 protons) lawfully supplies its salient properties such as boiling point and reactivity, so we might say that an atomic number of 79 comprises an essence that rigidly characterizes the kind gold. Insofar as essentialist kinds produce identical members with exceptionless properties, they are clearly impossible in psychiatry: there is simply too much variation. But an analog from medicine might be a disease stemming from a single aetiology found in all cases (Haslam, 2014). Huntington's chorea depends on a mutation of the *HTT* gene on chromosome-4, the presence of which produces a dependable aetiology that unfolds in a suitably law-like fashion. The mutated *HTT* gene approximates a microstructural essence that *is* the kind Huntington's chorea. However, many have persuasively argued that even this relaxed essentialism is far too strict an expectation for psychiatric disorders because their presentations are tremendously varied and multifarious (Murphy, 2006, pp. 132–149; Kendler & Schaffner, 2011). This is not to say that uncovering details about genetic or molecular contributions to mental disorder is wrongheaded (quite the opposite), but that it is unreasonable to expect this activity to resolve the apparent complexity of mental disorder to a neatly reductive, single (or simple) etiology *qua* essence. Common psychiatric disorders are not like Huntington's or HIV/AIDS; searching for *the* "gene" for depression is a wild goose chase. A good account of psychiatric disorders must instead reckon with the daunting complexity housed in the biological, psychological, and social sciences.

The usual tactic for explaining complex systems in mind and brain science is some form of bottom-up explanation like decomposition and localization (Bechtel & Richardson, 1993), whereby we break a system into its parts and specify how the nature and organization of those parts combine to create the whole. To explain the visual system, we parse it into several key sub-capacities realized by pieces of neural tissue and tell a causal story about how their coordinated activity generates the overall pattern we wish to explain.

These systems, however, are bedeviled by complexities that make explanation difficult. One is *compositional complexity*: the bits and pieces of biological systems are intricate, multifaceted, and put together in complicated ways (Mitchell, 2003). These systems – like an individual organism or ecological community – are made from many types of components that overlap, interact, and span organizational scales. Biological systems are thus multicomponent *and* multilevel.

If we stay strictly within traditional mind and brain science, some examples of putative components of psychopathology and their associated domains are serotonin transporter genes (genetics), dopamine deficits (biochemistry), hippocampus size (neurology), threat appraisal (cognitive neuroscience), shyness (traits), and learned helplessness (psychology). A difficulty is that many higher-order and less scientifically tractable phenomena also seem highly relevant to mental illness, and it is unclear how these mesh with the rest. For example, public humiliation (biography), self-concepts (beliefs), and social norms (sociocultural variables), often play indispensable roles in crafting mental illness, but we currently struggle to integrate them into our discussion of genes and neurochemicals. This entails problems of levels, explanation, and cause.

The second is *dynamic complexity*: a feature that arises when mutable, tightly connected components are sensitive and highly active (Mitchell, 2003). This kind of dynamism is endemic to much of biology. Zooming out, biological systems are dynamic because they adapt to changing environments in ways governed by the various forces of selection and adaptation. Zooming in, biological systems are often composed of malleable feedback loops that enable self-organization and regulation, which can change the operation and assembly of the system over time. Patterns of dynamic activity are often non-linear and potentially chaotic, where small perturbations to initial conditions produce erratic large-scale changes, like Edward Lorenz's "butterfly effect". This creates an epistemic difficulty because the behavior of a dynamically complex system resists prediction and explanation, even if the system unfolds deterministically and we can access all the right details about initial conditions.

Each type of complexity begets further problems for explanation. We can distinguish three types of systems: aggregative, component, and integrated (Levins, 1970; Bechtel & Richardson, 1993). We explain *aggregative systems* by simply combining the activities of parts, while *component systems* demand that we attend to parts plus their organisation. Rearrange a thousand elements in a gas cloud, and you haven't really changed the nature of the gas (aggregative). Rearrange a thousand genes in a sequence of DNA, however, and you've changed the nature of the DNA enormously. This is because DNA is highly organized (component). *Integrated systems*, meanwhile, are even trickier because both the parts and their organization dynamically interact with *each other* in ways that make bottom-up explanation lose its explanatory grip. There is a two-way street of dependency relations spanning up and down levels, so integrated systems demand bottom-up *and* top-down explanations. For example, individual bees in a honey colony behave differently depending on their gene expression, but their gene expression itself is regulated by social context, i.e., "the needs of the colony" (Mitchell, 2008, p. 29). Integrated systems are minimally decomposable and thus minimally friendly to bottom-up explanation. The brain with all its plasticity and feedback loops often behaves like an integrated system (Murphy, 2010).

2.2 Mechanistic Property Clusters

So, can we find kinds within the chaos? Kindhood is intimately tied to explanation (Cooper, 2005), so another route is to look for robust explanatory structures hiding within the mess that ground generalizations. If different things share this explanatory structure, they form a kind.

Following this, Kendler, Zachar and Craver (2011) argue that psychiatric disorders are mechanistic property cluster (MPC) kinds. The putative explanatory structures here are mechanisms that consist of entities, activities, and events that make a causal machine. Cause, although the term may not describe a single type of connection between events (Godfrey-

Smith, 2010), is often considered a *robust difference maker* that is indispensably responsible for the generation of a phenomenon (Woodward, 2003). Explaining a mechanism thus requires decomposition and localization into causally relevant components. This account pays a large debt to a view of kinds in the biological sciences advanced by Richard Boyd through several papers (e.g., Boyd, 1999, 2019). An MPC kind is made from (1) a cluster of properties, like signs and symptoms, that reliably co-occur in nature; and (2) a causal mechanism that explains the clustering. Consider stroke victims: patients can experience a constellation of different impairments (e.g., paralysis, confusion, blindness) none of which, individually or collectively, call for a stroke diagnosis. Here we have the familiar symptom variation. However, these symptoms arise and vary as a function of the location and degree of damage performed by an intracranial blood-clot. This is an underlying mechanism of vessel obstruction and neural death that causally accounts for different combinations of outwardly detectable symptoms. A stroke is an MPC kind.

The MPC account is attractive in psychiatry for two reasons. First, it accommodates the diverse experiences of patients who supposedly suffer the same affliction because characteristic properties like symptoms are expected to cluster imperfectly. Second, MPC kinds are incredibly flexible as to how the mechanism-cluster relationship can manifest[5]. A stroke involves a single lower-level event creating a higher-level suite of symptoms. But MPC kinds can also have complicated extensions spanning levels of organization, and symptoms themselves may play causal roles in a mechanism's activity. The account can (in principle) engross environmental variables and temporally distant risk factors into its picture of kindhood so long as we articulate them mechanistically, which involves sketching them as causally potent things that are locally connected and relevant to a psychopathology. Importantly, none of these factors are necessary or sufficient for kindhood. We instead have a family-resemblance-like structure where no members overlap on a strict set of features (Boyd, 1999, pp. 143–144), but all share important similarities to each other because relevant properties tend to non-accidentally hang together in nature. In this picture, mental disorder emerges from a complex of entities, activities, and events that criss-cross levels.

So, the ideal causal signature for the vindication project is some kind of stable, mutually reinforcing, multilevel mechanism that explains the clinical manifestations of a psychiatric syndrome. The more regularly the mechanism operates, the easier it will be to track.

2.3 Network Dynamics

I want to fine-tune the MPC account by synthesizing it with a vocabulary of network dynamics (e.g., Borsboom, 2017). Borsboom's theory emphasizes that symptoms are not inert consequences of disorders but properties with causal powers of their own. So, instead of mental illness emerging from some hidden "disorder" pulling secret strings, perhaps it simply consists of symptoms causing *each other* in a self-sustaining network. For example, someone with depression might have insomnia which causes fatigue which causes anxiety, which in turn causes insomnia. This pattern of symptom-symptom causation can be configured into a

[5] See Figures 1-3 in Kendler, Zachar and Craver (2011).

network structure where nodes represent symptoms and edges represent relations among symptoms. The clinical task is to map these networks in individuals and groups.

The emergence of mental illness depends on two dimensions: (1) the strength of connections between relevant nodes, and (2) the presence of environmental stressors. Importantly, Borsboom (2017, p. 9) defines mental health not as the absence of symptoms but as a stable equilibrium state to which a healthy system will return if disturbed. This means that mentally "healthy" and "ill" systems have the same general arrangement of causal connections between symptoms; what differs is the power of those connections and the pattern of activation that ensues. In a provocative environment (e.g., traumatising event), a healthy system may engage in patterns that bear the marks of mental illness (e.g., activations of low mood, hypersomnia, guilt) but the system will eventually return to a "normal" configuration after the stressful event loses potency. The system is resilient. One characteristic of a "disordered" system is that its activity becomes self-sustaining as nodes continue to activate *each other* even after the trigger has vanished.

This asymmetry in the transition between activation states is called *hysteresis*. Other concepts which will become important are *feedback loops* (especially positive ones) and *tipping points*, which describe the threshold at which a system will lock into a new stable pattern of activity.

Borsboom's account has two interesting things to say about the vindication project. First, and contrary to the alarmist picture presented in Section 1, perhaps our clinical intuitions embodied in the DSM have led us to a perfectly respectable picture of mental illness. The mistake is thinking we need to uncover some veiled "disorder" to understand the true nature of mental distress. While the DSM does not explicitly emphasize any causal connections between signs or symptoms, those relationships are often implicit in diagnostic criteria. Furthermore, clinicians in practice often deploy causal thinking about how relevant DSM-criteria interact within a particular patient (Borsboom, 2017, p. 9). So psychiatric classification might need revision, but it does not need revolution. Second, the causal signature that interests us is just the surface-level pattern of symptom interactions, so most disorders as currently defined simply vindicate themselves.

The problem is that Borsboom is betting the correct granularity at which to understand psychiatric disorders is at the level of symptoms, and nothing else. The account glosses over any detail about physical realization and information processing in minds and brains that may turn out to be genuinely explanatory over and above the activities of symptoms alone. It is premature to restrict a network picture of mental illness to one that only hooks up symptoms. Moreover, a symptom-symptom network is already a valid MPC kind: the causal mechanism is just the cluster of symptoms directly influencing each other[6].

But the importance of Borsboom's contribution lies in network dynamics. We can augment the MPC account by encoding the relevant entities, activities, and events as nodes; and their causal interactions as edges. MPC structures in minds and brains are dynamic and plastic so concepts of hysteresis, feedback loops, and tipping points are helpful in describing

[6] See Figure 2 in Kendler, Zachar and Craver (2011, p. 1174).

their behaviour over time. Thus, networks are useful representational tools that help explain the activities of mechanisms (Craver, 2017).

In summary, a good causal account of what mental disorders are, and thus what the vindication project ought to be looking for, is some stable MPC structure in thrall to network dynamics that together explain the clinical variables described by the DSM.

3. The Structure of Normal Fear

This section builds a psychological model of the "normal" fear system to represent it as an MPC kind. My goal here is to survey and integrate the (largely) cohesive research on fear in humans and apply a degree of idealization to bring some important elements into focus. It will form the foundation of an account of how putative changes to normal fear engenders phobia.

I will first briefly discuss some warring theories of emotions and argue that we can construct an account of fear without a strong commitment to any camp. I will then split the human fear system into three subsystems inspired by Kelly's (2011) treatment of disgust: the *acquisition subsystem*, the *execution subsystem*, and *downstream effects*. The functional leanings of this model are a consequence of the high level of abstraction required for those key sub-systems to crystallize and they do not contradict the mechanistic tenets of the MPC account because functional analyses can be interpreted as "mechanism sketches" (Piccinini & Craver, 2011).

3.1 What is Fear?

Evolutionarily speaking, the early mammalian environment harbored threats like predators and natural disasters that could strike quickly with little warning; fear and the behaviors of escape, avoidance, and aggression evolved as useful strategies for combatting danger (Horwitz & Wakefield, 2012). The minimum requirement for this ability is an adequate perceptual system to identify threats and a motor system to execute responses. These evolved in tandem. As increasing complexity gave way to more sophisticated systems, a rich inventory of tasks could be performed, such as threat evaluation, decision making, coordinated physical responses, and advanced forms of learning. Evolution tends to build sophisticated processes out of and on top of simpler ones, and so the neural circuits associated with threat detection and avoidance lie in deep brain regions with many direct connections between sensory inputs and motor outputs that avoid passageway through the cerebral cortex (Carr, 2015). Functionally, this means that fear can be triggered without conscious input. Fear also has a smoke-detector-like bias towards activation that favors false positives (Nesse, 2005), i.e., once bitten, twice shy. The unpleasant feelings of fear and anxiety evolved to lock in those patterns of behavior (LeDoux, 2015).

A fear event involves a suite of neurological, biological, psychological, and behavioral activities unfolding in a coordinated sequence. It can begin when sensory systems (often vision) detect a stimulus that subcortical circuits quickly judge as potentially threatening (Adolphs, 2013). This unleashes a flurry of chemical activity like the stress response, which dumps cortisol and adrenaline into the bloodstream, increasing heart rate, respiration, and mobilizing energy resources. This pathway is buried in deep brain regions also present in non-human animals and activates reflexively before conscious awareness of the stimulus kicks in (LeDoux, 2015). Psychologically, attention narrows, that characteristic *feeling* of fear arises, and the subject enters a motivational state ready to combat danger through freeze-flight-fight behaviors.

A fervent debate is raging in psychology around the nature and status of emotions like fear. The different camps mostly agree on the activities listed above, but disagree on details about their causal connections, temporal sequence, fundamentality, universality, emergence, function, and more. For example, Ekman's (1999) *affect program theory* holds that basic emotions like fear are universal programs installed into creatures by evolution that unfold and are experienced in the same way by everyone. The subjective feeling of fear is just another procedure in the program's code. LeDoux (2015) meanwhile thinks we must split an affect-program-like *survival circuit* that is indeed pre-programmed by evolution, from the *emotion* or subjective feeling of fear, which is individually constructed by a brain as it tries to understand swift changes in the environment and the body. More radically, Barrett (2017) thinks that all emotions are just functions of two dimensions of valence and arousal that are picked out and imbued with emotional meaning by subjects and society.

I argue that, regardless of which theory of emotions is correct, the subjective feeling of fear (a) is always unpleasant, and (b) plays a reinforcing role in learning. Feelings are not causally inert. Fear can thus be configured into an MPC structure if we constrain the relevance of the conscious feeling of fear to its functional causal role in the overall fear response. An MPC account of fear thus implicates the characteristic features of something like LeDoux's *survival circuit* (i.e., threat detection, psychological, physiological, behavioral responses) plus the functional relevance of the emotion associated with it, whatever that emotion *really* is. Together these form the "execution subsystem."

3.2 Fear Acquisition

Fears are selectively acquired (Seligman, 2016). Heights, spiders, and the dark are near-universal fear triggers that are considered either innately programmed or particularly sensitive to fear learning. Height is a persuasive example of an innate fear: visual cliff experiments suggest that infants are naturally afraid of heights (Horwitz & Wakefield, 2012, pp. 41-44). Meanwhile, compared to neutral stimuli, it is easy to condition human and non-human animals to fear things that likely played threatening roles in our evolutionary history, like snakes (Horwitz & Wakefield, 2012, p. 27). A strict demarcation of stimuli innateness vs. sensitivity is probably not warranted, as both qualities likely exist and interact for the same elicitors. The takeaway is that there appears to be a database of pre-programmed stimuli that are strongly associated with the fear response.

Fear can also be sculpted and strengthened through psychological learning mechanisms that imbue a neutral stimulus with an expectation of danger (LeDoux, 2015). Direct learning occurs via the first-hand experience of a stimulus. *Classical conditioning* involves the formation of non-conscious associations between stimulus and response, e.g., being attacked by a dog and forming an association between it and danger. This is strengthened through repetition. *Operant conditioning* occurs when choices lead to punishment or reward. Positive punishment occurs by *introducing* a noxious stimulus – approaching a dog and getting attacked, approaching another dog and becoming uncomfortable and distressed; and negative reinforcement by *removing* a noxious stimulus – escaping a dog attack, avoiding dog parks in the future. Punishment decreases dog-seeking behavior, and negative reinforcement increases dog-avoiding behavior. Classical conditioning plays a vital role in fear acquisition, while operant conditioning shapes fear maintenance. Both are highly relevant to phobia (Davey, 1992).

Observational learning and *social learning* are indirect mechanisms that occur when the sociocultural environment offers information that shapes fear (Horwitz & Wakefield, 2012; LeDoux, 2015). These could be in play after witnessing an accident (e.g., car crash), or via instruction, media broadcasting, or cultural fascinations (e.g., health issues, terrorism). Fears in this sense are socially contagious. The final learning mechanism is *extinction*, which will be discussed in Section 4.3.2.

3.3 Downstream Effects

Finally, fear activation produces lingering effects that can hook around and influence the acquisition and execution subsystems. An intense response – like a traumatizing experience – can lead to new avoidance patterns, habits, cognitive biases, and long-term sensitivity, that impact threat evaluation, anxiety levels, and broadcasts about the experience which provide information to others (LeDoux, 2015). The fear system is plastic and adaptable. Large changes usually require strong and/or repeated exposures to fear-triggering situations. However, smaller and innocuous activations of fear can also incite lingering changes, but these may be insignificant or uninteresting.

3.4 A Psychological Model of the Fear System

We can now construct a model of the fear system pitched at a relatively high level of abstraction (Figure 1). It is inspired by Kelly's (2011) treatment of disgust. The model consists of three functionally defined sub-systems which set the parameters for the fear response. Each box represents a distinct psychological process or piece of cognitive architecture, and each arrow represents a causal relation. In the *acquisition subsystem*, direct and indirect learning pathways are distinguished, with both contributing to a "database" of stimuli that are fear-triggering due to past experience. This database borders another containing "universal" elicitors that are innately specified. Dashed lines separating direct/indirect learning mechanisms and acquired/universal databases signify the potential for overlap. The *execution subsystem* is an expression of the structures and capacities that generate the fear response and are approximated by LeDoux's survival circuit. Threats are detected, evaluated, and then spark the typical cascade of activities. *Downstream effects* are a not structural component of the fear system but a representation of the lingering changes that occur after fear activation. Parallel to this pathway is the "fear emotion," as described by LeDoux, which is constructed out of, and in turn influences, the fear response[7]. This model emphasizes learning pathways that are created after the fear response which causally connect it, the downstream effects, and the fear emotion

[7] Note that this model can easily accommodate other theories of emotion. For example, the affect program would simply place the "fear emotion" inside the "execution subsystem" as some link within that causal chain. Barrett (2017) meanwhile would consider the emotion as existing as some function of valence and arousal (negative and high arousal) that is picked out as "fear" by the subject and their community. The emphasis is still on the causal contribution of the relevant valence/arousal scores to the way the emotional response unfolds and shapes the parameters for the next response.

to direct learning mechanisms. The fear system is a *feedback loop*. This captures the iterative and reinforcing character of normal fear which I believe to be a crucial component of fear dysfunction in phobia.

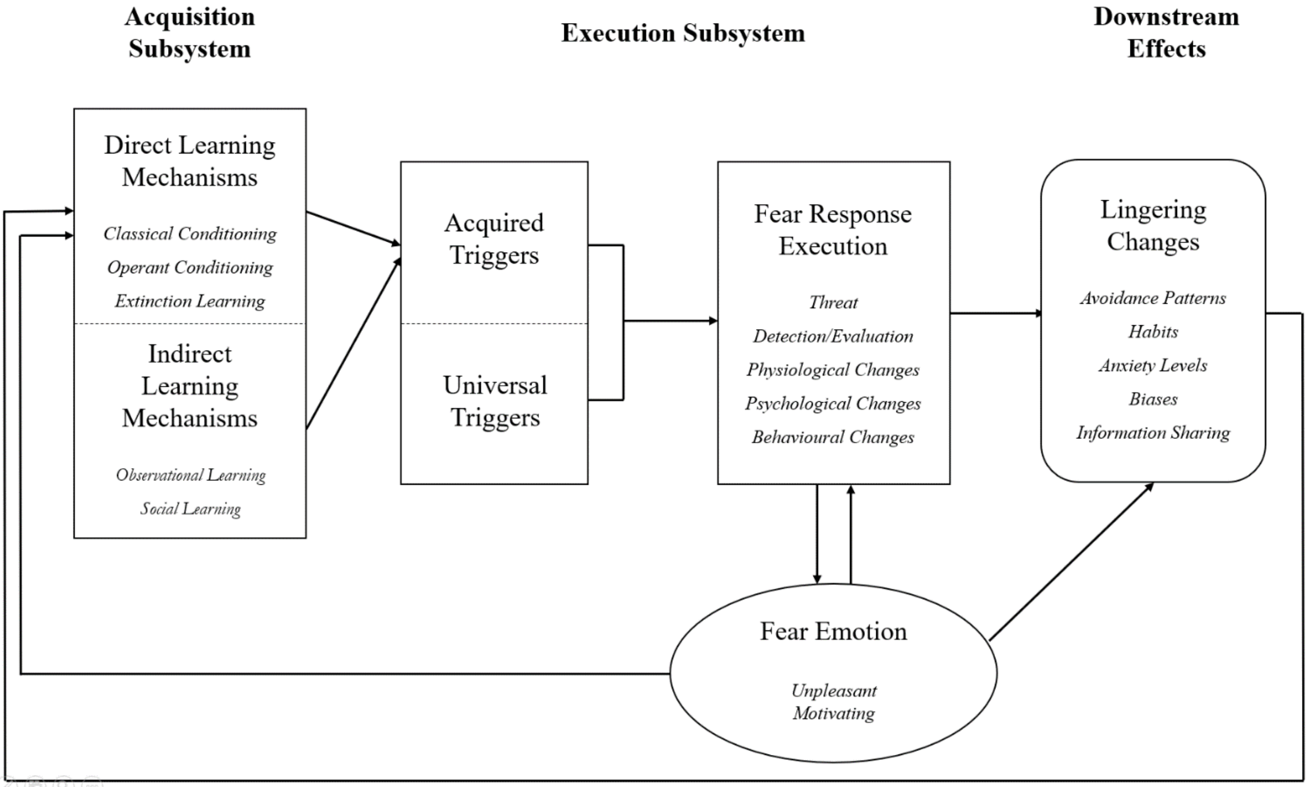

Figure 1. A functional model of the key subsystems and causal connection that underwrite the fear system

Finally, we can think of fear events as existing on a spectrum. At one end are the strong, simple, ephemeral episodes that involve a quick and dirty threat detection and reaction, e.g., jumping after finding a spider in your sandwich. This is the fear system acting neatly as evolution intended. At the other end are more abstract, uncertain, drawn-out, and conscious episodes of worry that involve some activity of this model being filtered through memory, identity, and complicated beliefs, e.g., anxiety in the weeks before a public-speaking event. This model has more explanatory purchase for simple fear episodes, and less for anxiousness and worry.

4. Specific Phobia is an Ideal Psychiatric Kind

Specific phobia is a remarkably simple way a mind can misfire. Evolution equipped us early on with a mechanism that forges a connection between a potentially dangerous stimulus and an inventory of defence responses, and specific phobia is a badly constructed stimulus-response relationship that secures overreaction to anodyne signals. In this section, I sketch a novel theory of phobia in which it is characterized by a general pattern of fear dysregulation. The task is to take the clinical variables described by the DSM-5 and hook them up to a causal account of *changes* to the fear response that explains why those properties cluster together. Although speculative, it is empirically informed and matches many intuitions about anxiety disorders. Normal fear and phobia are two different states of a common plastic system that are distinguished by the nature and strength of causal connections between a wide array of multi-level interactions. In this, I will argue that specific phobia is an ideal MPC kind of psychopathology.

4.1 Specific Phobia in the DSM-5

The DSM-5 houses specific phobia under the umbrella of "anxiety disorders" and they involve an excessive and unreasonable fear of a circumscribed object, situation, or event (APA, 2013). The triggers can be real or imagined. Specific phobia is distinct from social anxiety disorder (known previously as social phobia), the fear of social settings, and agoraphobia, the fear of situations that can trigger panic attacks. Phobia sub-types are codified by different sorts of fear-inducing stimuli:

Animal – e.g., spiders, dogs.

Natural Environment – e.g., heights, water.

Blood-Injection Injury – e.g., needles, surgeries.

Situational – e.g., airplanes, elevators

Other – e.g., choking, dentists, dolls.

Key features are: (i) that the phobic stimulus regularly incites fear/anxiety, (ii) this fear/anxiety is disproportionate to the actual danger presented by the stimulus, and (iii) the fear/anxiety prompts significant avoidance patterns that disrupt normal functioning to levels of clinical significance. Specific phobia may be the most common mental illness in the world (lifetime prevalence estimated at 7-11%), along with major depressive disorder and social anxiety disorder (APA, 2013; Bandelow & Michaelis, 2015). Phobia is often comorbid with depression and, unsurprisingly, other anxiety disorders. Some patients recall a traumatic experience that triggered their phobia, while others cannot; there are many pathways to phobia (APA, 2013).

Recall that are three striking characteristics of specific phobia which together suggest it is an ideal mental disorder explicable by the MPC account. First, specific phobias are cross-culturally stable. The expression and phenomenology of phobia are consistent across time and place which suggests we are dealing with a common underlying state-of-affairs that spans cultural boundaries. There are slight differences across demographics and larger differences across the content of phobias (i.e., *what* is feared), which are likely culturally inherited (Marques, Robinaugh, LeBlanc & Hinton, 2011). However, this pancultural consistency is very rare for a mental disorder. Second, we have a persuasive scientific account of fear (Section 3). Fear is evolutionarily ancient and is triggered automatically with minimal conscious input, rendering it a more scientifically tractable psychological process. In terms of a psychological system breaking down, a cohesive account of fear will ground a cohesive account of fear dysfunction. Finally, phobia is regularly cured through exposure therapy (Choy, Fyer & Lipstitz, 2007). It is extraordinarily rare for a mental disorder to be undone by a single non-individualized treatment that appears to act directly (even surgically) on the underlying problem, in almost all cases. In fact, this may be the *only* example. Interventionist accounts of causation (e.g., Woodward, 2003) would say that this striking success permits us to think that the activities exposure therapy intervenes upon are *the* causal agents responsible for creating and maintaining phobia.

4.2 What Makes a Phobia?

There are two general routes to explaining phobia, but neither path has led to a rigorous philosophical or scientific treatment. The first argues that phobics suffer from an evolutionary mismatch between a once adaptive trait and a novel environment (e.g., Nesse, 1999; Horwitz & Wakefield, 2012, pp. 51–79), the other because broken mechanisms render phobics unduly anxious (Murphy, 2006, pp. 281–307). The mismatch hypothesis observes first that anxiety displays normal phenotypic variation, leaving those on the sensitive end of the distribution naturally more anxious, and second that many fears involve stimuli that likely played threatening roles in our evolutionary history, like snakes and the dark (Section 3.2). On this account, a snake phobia is an evolutionary hangover that arises in part because modern medicine and infrastructure have neutralized the threat once posed by snakes; an inherited fear is mismatching with present-day circumstances. Mismatching provides insight as an ultimate explanation of how one camp of stimuli can ground phobia (“universal triggers”) but does little to explain phobias of evolutionarily irrelevant fears, like dolls (pediophobia). Evolution bequeathed us with an adaptable machine that can learn (and crucially, unlearn) to fear almost anything. We must, therefore, attend to proximal mechanisms and expect to find some putative breakdown-like difference in phobia that can engender cyclic patterns of rigid learning and disproportionate reaction to a wide array of stimuli. A complete picture of phobia requires some integration of evolutionary and breakdown explanations, but the emphasis should be placed on the latter.

The way into phobias is to observe that what is unreasonable about them is not necessarily the content or excessiveness of fear, but rather the *formation* and *maintenance* of fear and its character of *hijacking* a person’s life through distress and avoidance. Phobia obstructs desires, impedes life-goals, and prevents flourishing. Phobia spawns from a general failure of fear regulation.

Consider this hypothetical. Snakes are panculturally fear-triggering, and so a neurotypical person, call them Andy, will likely fear a snake that appears suddenly at a family barbeque. The fear pathway leads from detection to defensive behaviors, and so once activated, Andy will be pushed to freeze immediately and then either run away or attack, pending some quick evaluations. But this probably won’t happen. Instead, Andy will *down-regulate* and thus override their fear to execute goal-oriented tasks like consciously assessing the danger, warning others, and getting children to safety. A person with a snake phobia, call them Blake, will likely act in accordance with the fear response and flee even if such tasks are valued (like protecting children). The fear is overwhelming. Furthermore, the experience may have a potent effect on Blake’s learning and produce future avoidance of barbeques. Andy, albeit shaken, can regulate their learning and not let the event dictate future behavior.

Fear dysregulation, in this broad sense, characterizes the inability of the plastic system to control the unfolding of the fear response. There are two general types of regulation strategies. *Automatic regulation* involves unconscious processes that regulate via in-built biological mechanisms, such as homeostatic neurochemical relationships. *Effortful regulation* involves a person consciously trying to control an emotional response as it occurs (Cisler & Olatunji, 2012) such as stifling laughter or trying to conjure up gratitude for a disappointing Christmas gift. The hallmark feature of phobia is that a powerful stimulus-response association obtains despite attempts to regulate a response either automatically or effortfully, over both immediate and extended timescales.

4.3 Two Mechanisms of Dysregulation

The MPC view and thus the vindication project holds that we should be looking for multi-level causal mechanisms to explain complex systems. Here, I will zoom in on two sub-systems, pitched at different levels, that empirical evidence suggests are dysregulated in phobics. One is a homeostatic neurochemical relationship, the other an information processing mechanism.

4.3.1 Regulation of the HPA Stress Axis

The "stress response" involves the activation of two systems: the sympathetic nervous system and the hypothalamus-pituitary-adrenal (HPA) axis (LeDoux, 2015). The sympathetic nervous system is activated immediately after non-conscious threat detection and works via autonomic nerves to release epinephrine into the bloodstream, which increases heart rate, respiration, and alertness. This epinephrine surge quickly subsides, and the HPA axis kicks in to release the stress hormone cortisol which mobilizes resources for fight-flight-freeze responses; in short, it keeps the sympathetic nervous system revved up.

Crucially, cortisol release is modulated by a push-pull relationship between the amygdala and hippocampus (Juruena, Cleare & Pariante, 2004) [**Figure 2**][8]. The hippocampus controls the release of cortisol by inhibiting the amygdala as higher levels of cortisol flood the bloodstream. The connection forms a negative feedback cycle that counterbalances the excitatory zeal of the amygdala to return cortisol to homeostatic levels.

Now, many neurological studies indicate that anxious people have hyperresponsive amygdalas and difficulties regulating fear (Grupe & Nitschke, 2013). In one study, spider phobics were asked to effortfully up and down-regulate their fear in response to pictures of (a) spiders, (b) generally aversive stimuli, and (c) neutral stimuli. They found that both automatic and effortful regulation were impaired during exposure to spiders but not to the generally aversive or neutral stimuli, suggesting selective impairment towards the phobic stimulus (Hermann, Schäfer, Walter, Stark & Schienle, 2009). A strong pattern of change to limbic and cortical activity could plausibly destabilize the modulation of cortisol release. I am not suggesting that changes to limbic activity is a fundamental explanation of phobia, but showing a specific example of fear dysregulation realized in low-level biological mechanisms that might play a role in a complicated MPC structure.

[8] Diagram is a simplified version of Figure 1, Juruena, Cleare, & Pariante (2004, p. 191).

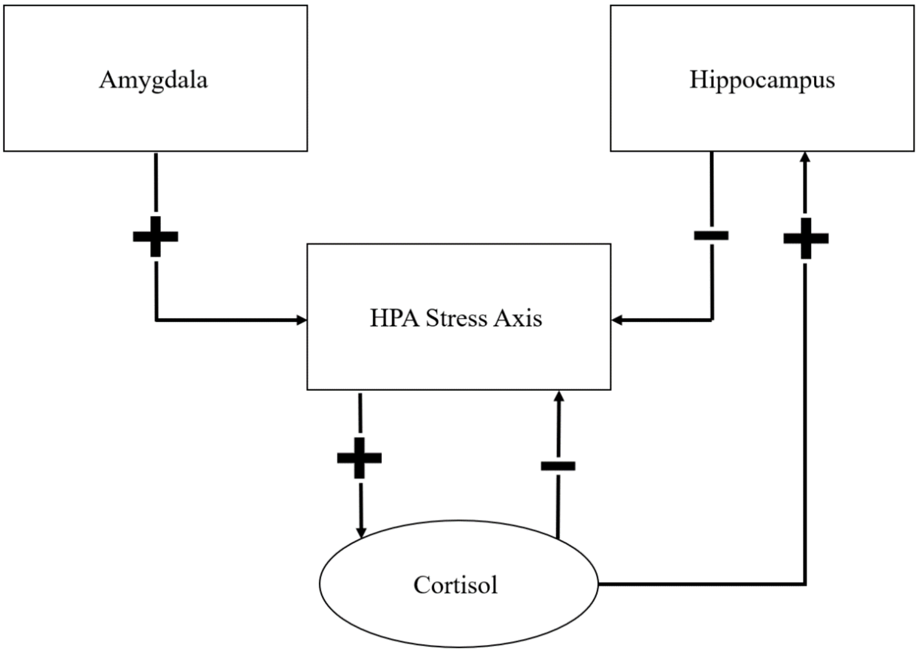

Figure 2. The push-pull relationship between the amygdala and hippocampus regulating the release of cortisol by the HPA stress axis.

4.3.2 Extinction Learning and Exposure Therapy

I left extinction learning out of Section 3.2 to address it here, but it should be considered alongside the "direct learning mechanisms" in the structure of normal fear. *Extinction learning* opposes those processes by decreasing the behavior shaped by a stimulus-response relationship across trials (LeDoux, 2015). If a rat has been fear-conditioned to avoid a particular area because it invariably leads to an electric shock, then this pattern of avoidance can be extinguished by allowing the rat to enter the area without consequence. This association, as in operant and classical conditioning, is strengthened over trials. However, extinction learning does not equal "unlearning" or "forgetting" but is often interpreted as new learning that reflects a gradual dissociation between the initial stimulus-response relationship. In the context of normal fear, extinction works to counterbalance the reinforcing effects of negative fear experiences on future behaviors.

We should also note that cognitive activities like attention and bias influence learning. For example, a belief that airplanes are unsafe might ground a confirmation bias that narrows attention on frightening turbulence instead of the wealth of other information that flying is harmless. This impairs extinction learning.

Exposure therapy, as mentioned previously, is a highly successful treatment of phobias that is almost always used and often leads to cures[9]. It involves a patient gradually facing their

[9] It should be noted that exposure therapy, unsurprisingly, still has limitations. First, biographical factors and comorbidities (for example, OCD) can undermine exposure therapy because obsessive thoughts continually generate new phobias. Second, exposure therapy cannot overcome traditional problems with extinction learning, for example, context dependence, spontaneous recovery, and reinstatement (Quirk & Mueller, 2008).

fear in a real or imagined context to extinguish the stimulus-fear relationship that drives a phobia (LeDoux, 2015). A snake phobic might begin by drawing a picture of a snake, then observing one through a glass window, and finally, after multiple trials that steadily increase stimulus potency, enter a room with a snake. The goal is to slowly and systematically introduce snake-like stimuli to initiate extinction. The patient is importantly not passive. Even moderate stimuli evoke intense responses, and so the patient must be motivated and courageous for the treatment to be effective.

This is, essentially, a vehicle of fear regulation. I think we can picture it in two ways. First, it supplies a potent source of extinction information that might have, in a typical person, prevented the non-threatening stimulus from hijacking fear in the first place. It causally treats phobia by fuelling extinction learning that had previously failed. This dimension of exposure therapy acts on non-conscious conditioning mechanisms of the sort successfully studied in animal models (e.g., LeDoux, 2015, pp. 282–320). Second, exposure therapy constructs a safe context that allows other conscious regulation strategies to become effective. The mild stimulus and clinical setting allow a patient to *practice* effortful regulation, providing an opportunity to correct bad conscious and automatic control systems. Thus, we can consider "extinction learning", or the broad mix of activities required to achieve this, as another mechanism of fear regulation that cannot take place in phobia without clinical assistance.

4.4 An MPC Account of Specific Phobia

The last section drew on empirical evidence that phobia is associated with an impaired ability to regulate fear at different levels of organization. Here, I want to integrate the various parts of this paper to show that phobia constitutes an MPC kind. This requires us to take the clinical features recorded by the DSM and anchor them in some causal-mechanical destructive process that explains why those features arise and maintain. I argue that phobia is characterized by a broad pattern of multi-level fear dysregulation that disrupts the typical activity of the fear system and locks it into a stable, inflexible pattern of intense reactions and powerful avoidance.

First, how can we make sense of the many causal processes that seem to contribute to phobia? I have argued what we should not expect a simple reductive explanation nor a single aetiological pathway; instead, we must pitch an account that hooks up multiple causal domains. It should be clear that there is no keystone strategy or single regulation technique required to go wrong in phobia.

I hold that phobia involves a series of causes that are neither necessary nor sufficient, which may or may not interact, producing an *underlying state* – fear dysregulation – that accounts for the clinical features of the DSM-5. This is represented in Figure 3 (Kendler, Zahcar & Craver, 2011, p. 1147). The relevant causes might involve a hyperresponsive amygdala, traumatizing experience, diminished capacity for extinction learning, cognitive bias, or any combination, but the motley cocktail of causes nevertheless creates a stable phenomenon reliably tracked by DSM.

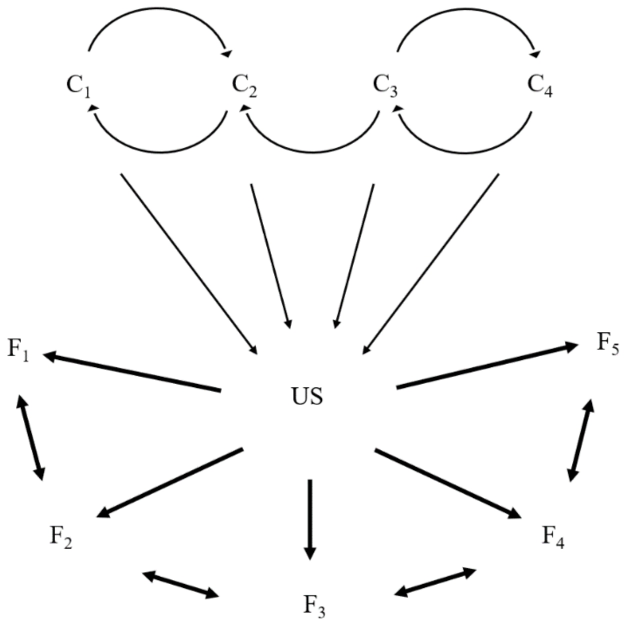

Figure 3. Several interacting causes (C_1-C_4) create an underlying state (US). The US produces a cluster of clinical features (F_1-F_5) that may or may not interact with each other. These are picked out by the diagnostic criteria of the DSM-5.

Second, we must appreciate that fear is iterative with each instance of the fear response hooking back around and influencing the parameters for the next response. Fear is loopy. It is thus especially vulnerable to positive feedback cycles (Section 3.4), particularly when considering the evolutionary benefits of a "smoke-detector" like bias toward false-positives (Nesse, 2005). Fear depends on a dynamic system whose triggers, execution mechanisms, and downstream effects are constantly readjusting. These changes are usually minuscule; normal fear is steady.

However, a single cause or collection of causes can begin to engender dysregulated activity that positive feedback amplifies and cements. Dysregulation is not so much a cause or a consequence but a *catalyst* that breaks the cycle of normal fear and locks it into a new pattern of intense reaction and avoidance that becomes inflexible and resists regulation. Phobia spirals out of normal fear. As this spiral unravels, other properties characteristic of phobia might appear (e.g., cognitive biases, physiological changes to the amygdala) and participate in solidifying the cycle.

We thus have two broad configurations or states of the fear system: one characterized by flexible but stable connections regulated by homeostatic mechanisms, and another characterized by inflexible and entrenched connections that resist regulatory control. Each state corresponds to its own MPC kind, and a particular system can shift between the two. Phobia endures because dysregulated activity exerts a gravitational-like force on the fear system that pulls and traps it into a new stable state, like a marble rolling into a basin (*tipping point*). If a system of risk-factors is exposed to a traumatizing environment the phobic state can arise and then self-sustain (*hysteresis*). A clinical intervention like exposure therapy is required to break this stable cycle and restore regulatory control.

Thus, phobia is caused by some permutation of the relevant parts of a causal-mechanical system misfiring to varying degrees in concert. Despite the many possible pathways and configurations of aberrant components, all roads lead to a common state that endures with remarkable stability. This explains the three striking features of phobias described in Section 4.1. So, a consequence of the structure of the fear system and its connections to everything else is that it is vulnerable to falling into this regular pattern of intense activation and powerful avoidance, and several different antecedent and constitutive causes might be involved. These nonetheless generate a reliable cluster of self-sustaining features that are detected by the DSM-5. If this dysregulation secures an inappropriate fear response towards a circumscribed stimulus, it creates a specific phobia.

5 Discussion and Concluding Remarks

I began this paper by noting an overarching feeling of pessimism towards the DSM's ability to guide psychiatry to scientifically defensible kinds of psychopathology. This was countered by optimism in a new program of research and taxonomic revision that emphasizes neural, biological, and psychological causal mechanisms. Although this program abandons DSM disorders as research targets, a strand of it aligns with Murphy's (2014) vindication project: we might validate a DSM diagnosis and promote it to a scientifically defensible psychopathology if we can discover a vindicating "causal signature" – some stable, multi-level causal mechanism – that explains why the syndromic features of a putative disorder appear and persist. This project will help us determine which of our current mental disorders are trustworthy, and which require revision.

I then argued that specific phobia is our best example of a causally vindicated (or vindicate-able) psychopathology because it fits Kendler, Zachar, and Craver's (2011) MPC account. This means we have a legitimate, "gold-standard" psychiatric diagnosis that looks as medical as anything. Specific phobias operate panculturally, arise from the tractable dysfunction of a (relatively) simple and primitive psychological capacity, and are successfully intervened upon by a single treatment. To close, I will reflect on the model and speculate on exactly how optimistic we ought to be about the success of the vindication project.

This model is intentionally restricted to *specific* phobia and not agoraphobia and social anxiety disorder, its taxonomic neighbors. I mentioned in Section 3.4 that a particular fear event lies on a continuum between quick and dirty reactions and extended periods of existential worry. The interestingness of specific phobia hinges on the misfiring of a basic psychological mechanism that, while undeniably "mental", happens to avoid conscious input. I suspect this partially explains why specific phobia is so stable a disorder. Things change as we move to agoraphobia and social anxiety disorder because these push towards the opposite end of the spectrum. For example, social anxiety disorder is not a simple fear of an object but a general and diffuse dread of public humiliation and judgement. Explaining it involves attending to slipperier concepts like self-esteem, rationality, identity, and beliefs – it exhibits a more integrated and top-down character which limits the explanatory purchase of the present model (Section 2.1), and further opens the disorder to sociocultural shaping. Factors like belief and rationality are certainly not absent in specific phobia, but appear muted enough to keep the disorder simple and stable.

This leads us to ask: does the MPC account ask too much of psychiatric disorders? Recall that the MPC account is attractive due to its ability *in principle* to absorb all the entities, activities, and events relevant to psychopathology. But how feasible is this in practice? The effectiveness of the vindication project depends on our ability to anchor the syndromic features of disorder in a causal-mechanical story, but it is not yet clear how we might bundle together details about genes and brain circuits with thoughts, beliefs, and lived experience. Perhaps they are immiscible. Or, perhaps we will eventually reduce beliefs and so on to brain circuits, or eliminate the concepts altogether. The burden is on MPC proponents to articulate how exactly these relate in a causal machine, and why the MPC account is superior to alternative explanation strategies that don't employ causal mechanisms at all.

For now, it seems that specific phobia is a special case. Specific phobia holds a unique place in our picture of mental disorder that has so far been missed. It is a "gold standard" mental disorder; an ideal psychiatric kind.

Acknowledgements

First and foremost, I would like to thank Dominic Murphy and Peter Godfrey-Smith for their helpful comments and supervision, along with Daniel Kelly and Georg Repnikov for their insightful feedback on an earlier draft. This paper has also benefitted from informal input by Maureen O'Malley, Kristian Camilleri, Paul Griffiths, Rob Wilson, Carrie Figdor, and Gemma Lucy Smart. I am also indebted to the Association for the Advancement of Philosophy and Psychiatry for their help and support.

Declaration of Conflicting Interests

The author declares no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research is supported by an Australian Government Research Training Program (RTP) Scholarship.

References

Adolphs, R. (2013). The Biology of Fear. *Current Biology*, *23*(2), R79–R93.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Bandelow, B., & Michaelis, S. (2015). Epidemiology of anxiety disorders in the 21st century. *Dialogues in Clinical Neuroscience*, *17*(3), 327–335.

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain.* Houghton Mifflin Harcourt.

Bechtel, W., & Richardson, R. C. (1993). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton University Press.

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375

Boyd, R. (1999). Homeostasis, Species and Higher Taxa. In R. A. Wilson (Ed.), *Species: New Interdisciplinary Essays* (pp. 141–185). MIT Press.

Boyd, R. (2019). Rethinking natural kinds, reference and truth: Towards more correspondence with reality, not less. *Synthese*, 1–41. https://doi.org/10.1007/s11229-019-02138-4

Carr, J. (2015). *I'll take the low road: The evolutionary underpinnings of visually triggered fear* (Vol. 9). https://doi.org/10.3389/fnins.2015.00414

Choy, Y., Fyer, A. J., & Lipsitz, J. D. (2007). Treatment of specific phobia in adults. *Clinical Psychology Review*, *27*(3), 266–286. https://doi.org/10.1016/j.cpr.2006.10.002

Cisler, J. M., & Olatunji, B. O. (2012). Emotion Regulation and Anxiety Disorders. *Current Psychiatry Reports*, *14*(3), 182–187. https://doi.org/10.1007/s11920-012-0262-2

Cooper, R. V. (2005). *Classifying madness: A philosophical examination of the diagnostic and statistical manual of mental disorders*. Springer.

Davey, G. C. L. (1992). Classical conditioning and the acquisition of human fears and phobias: A review and synthesis of the literature. *Advances in Behaviour Research and Therapy*, *14*(1), 29–66. https://doi.org/10.1016/0146-6402(92)90010-L

Ekman, P. (1999). Basic Emotions. In T. Dalgleish & M. J. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 45–60). John Wiley & Sons, Ltd. https://doi.org/10.1002/0470013494.ch3

Godfrey-Smith, P. (2010). Causal Pluralism. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *Oxford Handbook of Causation* (pp. 326–337). Oxford University Press.

Greenberg, G. (2013). *Book of Woe: The DSM and the Unmaking of Psychiatry*. Scribe Publications.

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience*, *14*(7), 488–501. https://doi.org/10.1038/nrn3524

Haslam, N. (2014). Natural Kinds in Psychiatry: Conceptually Implausible, Empirically Questionable, and Stigmatizing. In H. Kincaid & J. A. Sullivan (Eds.), *Classifying Psychopathology: Mental Kinds and Natural Kinds*. MIT Press.

Hermann, A., Schäfer, A., Walter, B., Stark, R., Vaitl, D., & Schienle, A. (2009). Emotion regulation in spider phobia: Role of the medial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, *4*(3), 257–267. PMC. https://doi.org/10.1093/scan/nsp013

Horwitz, Allan V., & Wakefield, J. C. (2007). *The loss of sadness: How psychiatry transformed normal sorrow into depressive disorder*. Oxford University Press.

Horwitz, A.V., & Wakefield, J. C. (2012). *All We Have to Fear: Psychiatry's Transformation of Natural Anxieties Into Mental Disorders*. Oxford University Press, USA. https://books.google.com.au/books?id=Hp6_p4tgV9kC

Insel, T. (2013, April 29). *Transforming Diagnosis: Post by Former NIMH Director Thomas Insel*. https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2013/transforming-diagnosis.shtml

Juruena, M. F., Cleare, A. J., & Pariante, C. M. (2004). The hypothalamic pituitary adrenal axis, glucocorticoid receptor function and relevance to depression. *Revista Brasileira de Psiquiatria*, *26*(3), 189–201.

Kelly, D. (2011). *Yuck! The Nature and Moral Significance of Disgust*. The MIT Press.

Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders? *Psychological Medicine*, *41*(6), 1143–1150. https://doi.org/10.1017/S0033291710001844

Kenneth S Kendler, & Schaffner, K. F. (2011). The Dopamine Hypothesis of Schizophrenia: An Historical and Philosophical Analysis. *Philosophy, Psychiatry, & Psychology*, *18*(1), 41–63. https://doi.org/10.1353/ppp.2011.0005

Kleinman, A. (1987). Anthropology and psychiatry. The role of culture in cross-cultural research on illness. *British Journal of Psychiatry*, *151*, 447–454.

LeDoux, J. (2015). *Anxious: Using the Brain to Understand and Treat Fear and Anxiety*. Viking (Penguin Random House).

Levins, R. (1970). Complex Systems. In C. H. Waddington (Ed.), *Towards a Theoretical Biology* (pp. 73–88). Edinburgh: University Press.

Marques, L., Robinaugh, D. J., LeBlanc, N. J., & Hinton, D. (2011). Cross-cultural variations in the prevalence and presentation of anxiety disorders. *Expert Review of Neurotherapeutics, 11*(2), 313–322. https://doi.org/10.1586/ern.10.122

Mitchell, S. (2008). Explaining Complex Behaviour. In K. S. Kendler & J. Parnas (Eds.), *Philosophical Issues in Psychiatry: Explanation, Phenomenology, and Nosology* (pp. 19–38). Johns Hopkins University Press.

Murphy, D. (2006). *Psychiatry in the Scientific Image*. MIT Press.

Murphy, D. (2010). *Complex Mental Disorders: Representation, Stability and Explanation*. *6*(1), 15.

Murphy, D. (2014). Natural Kinds in Folk Psychology and in Psychiatry. In H. Kincaid & J. A. Sullivan (Eds.), *Classiyfing Psychopathology: Mental Kinds and Natural Kinds* (pp. 105–122). MIT Press.

Nesse, R. M. (1999). Testing evolutionary hypotheses about mental disorders. In S. Stearns (Ed.), *Evolutionary Medicine* (pp. 260–266). Oxford University Press.

Nesse, R. M. (2005). Natural selection and the regulation of defenses. *Evolution and Human Behavior, 26*(1), 88–105. https://doi.org/10.1016/j.evolhumbehav.2004.08.002

Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese, 183*(3), 283–311.

Quirk, G. J., & Mueller, D. (2008). Neural Mechanisms of Extinction Learning and Retrieval. *Neuropsychopharmacology, 33*(1), 56–72. https://doi.org/10.1038/sj.npp.1301555

Seligman, M. E. P. (2016). Phobias and Preparedness. *Behavior Therapy, 47*(5), 577–584. https://doi.org/10.1016/j.beth.2016.08.006

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.