Obsessive-compulsive disorder as a disorder of attention.

Abstract: It is often held that obsessive-compulsive disorder is caused (in part) by distinctive personality traits and belief biases of sufferers. But while there is good evidence for a correlation between OCD and the cognitive traits these models emphasize, the correlation is by no means perfect: a substantial number of sufferers do not manifest these traits. In this paper, I propose a predictive coding account of the disorder, which aims to explain both the symptoms and the cognitive traits on which the metacognitive models turn. On this account, OCD centrally involves heightened and dysfunctionally focused attention to sensory and motor representations that are normally unattended. Because these sensory and motor representations have contents that predict catastrophic outcomes, patients are disposed to engage in behaviors and mental rituals designed to forestall these outcomes or to produce more precise information. The same representations also may cause the cognitive traits characteristic of sufferers: thus the same dysfunction gives rise to both the metacognitive dispositions correlated with OCD and to its symptoms.

Obsessive-compulsive disorder is a relatively common disorder which is responsible for a great deal of distress in sufferers and in many cases is highly disabling (Abramowitz, Taylor & McKay 2009; Veale & Roberts 2014). In most cases, OCD patients possess good insight into the pathological nature of their own cognition. They might compulsively perform rituals which, as they recognize, have no realistic connection to the events they thereby seemingly attempt to forestall, or wash their hands dozens of times a day, though they understand that once is enough (or more than enough). Sufferers report feeling compelled to perform rituals or being tormented by obsessive thoughts.

Distinctive cognitive and metacognitive traits seem to play a role in the etiology of OCD. On the most influential cognitive model, OCD begins with dysfunctional *appraisals* of unwanted thoughts (Salkovskis 1985; Wells 2000). The unwanted thoughts themselves are typically thoughts of engaging in behaviors that are (to the patient) unacceptable: defiling a crucifix, shouting an obscenity, swerving across three lanes of traffic while driving, strangling a loved one. Thoughts with contents like these are not confined to patients, however: they are in fact common in the non-clinical population (Gibbs 1996). OCD arises, on this model, not from the unwanted thoughts themselves but from their appraisal as predictive of behavior or as a basis for assessment of the agent. Compulsions arise as a way of preventing the behavior or the thoughts. In other words, compulsive behavior aims to elicit evidence that the patient is not engaged in some catastrophic behavior or is not normatively reprehensible.

While the cognitive model is influential, it seems unable to account for all cases. Although many OCD patients have the psychological characteristics predicted by the model, many do not. In this paper, I will argue that these cognitive peculiarities typically seen in OCD are not basic to it, though they may play

I am very grateful to two reviewers for *Mind & Language* for very helpful comments on this paper. In particular, I owe a debt to reviewer 1, who has opted to be identified as Karl Friston. He made extensive and detailed suggestions for improving the paper, most of which I have adopted. I am also grateful to an audience at the University of Wellington, New Zealand.

a causal role in those who exhibit them. Instead, heightened awareness of the contents of sensory and motor representations give rise to both the disorder *and* to these cognitions and metacognitions. We should not expect to see these psychological peculiarities in all OCD sufferers; we should instead expect to see peculiarities in the allocation of attention. I will situate this account within the broad framework provided by predictive coding accounts of cognition and action.

My argument for a predictive coding account of OCD fit comfortably with related accounts of other neuropsychiatric conditions; ranging from autism (Haker, Schneebeli & Stephan 2016), to psychosis (Corlett, Frith & Fletcher 2009; Adams et al. 2013), to hysterical (functional medical) symptoms (Edwards et al. 2012). The underlying premise here is that nearly all psychiatric symptoms can be understood in terms of false inference. The emerging picture is that aberrant beliefs may reflect a failure to encode the confidence or precision that should be assigned to sensory evidence. In other words, the primary pathology lies not in the way we explain the content of our perception but in the way we balance competing prior beliefs and sensory input. Mathematically, precision corresponds to the inverse variance or uncertainty associated with a signal or probability distribution. Physiologically, it is thought to correspond in the neuromodulatory gain of neuronal populations encoding beliefs about the world. As we will see later, the role of precision is central to my argument, and to its formulation in terms of attentional selection and action selection.

In the first section of this paper, I provide further details about the nature of OCD and sketch the theory that is currently the most influential. Doing so will lay bare both the facts that require explanation and the limitations of the currently best supported explanation of them. In the second section, I will introduce the predictive coding framework and an existing proposal for explaining OCD within it (Moore 2015). On this proposal, OCD arises when agents assign probabilities to counterfactual narratives dysfunctionally. While the central insight of this account is plausible, it appears to commit us to claims about the dispositions of OCD patients which are not supported by the available data. I will suggest that we need a different account of how sufferers come to assign probabilities dysfunctionally.

On the alternative account I will develop, both the symptoms of OCD and the cognitive and metacognitive dispositions typical of (but not invariably found) in sufferers arise from overly precise pushmi-pullyu sensory and motor representations, where a 'pushmi-pullyu representation' is both descriptive and directive, simultaneously depicting of a state of affairs and generating a set of motor representations aimed at bringing that state about (Millikan 1995). I will then argue that this account explains the (meta)cognitive traits typically seen in OCD. Appraisal style may stabilize OCD and make it more intractable, but it is not a necessary condition of the disorder, and may itself arise from the more basic dysfunction which I propose is the cause of OCD.

1. *Obsessive Compulsive Disorder: Symptoms and Explanations*

OCD is a symptomatically heterogeneous condition. It is standard to identify five principal symptom types (Abramowitz, Taylor & McKay 2009):

1. Obsessions about causing or failing to prevent harm and associated checking and reassurance seeking behavior;
2. Symmetry obsessions and associated ordering and counting rituals;
3. Contamination obsessions and associated washing and cleaning rituals;
4. Personally repugnant obsessions (concerning sex, religion or violence);
5. Hoarding and collecting impulses.[1]

Obsessions are recurrent and persistent intrusive thoughts, which can cause those who experience them great anxiety. The occurrence of these thoughts has sometimes been held to be the key to explaining OCD, but in fact their occurrence does not distinguish the clinical population from the non-clinical (Gibbs 1996). The currently most influential model instead holds that the dysfunctional appraisals of these thoughts is the key to the explanation (this model is, therefore, a two-factor model, structurally similar to two-factor models of delusion).

An appraisal is a metacognition: a thought about a thought. According to the influential model developed by Salkovskis (1985), intrusive thoughts become obsessions when they are catastrophically misinterpreted. Salkovskis suggested that an inflated sense of responsibility for preventing harms explains this misinterpretation. Later work postulated – and provided evidence for – other traits that may cause dysfunctional appraisals. The Obsessive Compulsive Cognitions Work Group (OCCWG) identified three domains of dysfunctional beliefs: (1) overestimation of threats and an inflated sense of responsibility; (2) the importance of a need to control thoughts; and (3) perfectionism and intolerance of uncertainty (OCCWG 2003; 2005). The second domain is correlated with or causes what has come to be called thought-action fusion (TAF), which itself comes in two main forms: a moralized form, in which the person is disposed to think that having a thought is morally revelatory (the thought of killing her child reveals to a woman that she is a bad mother, for instance), and a predictive form, in which having a thought entails a raised likelihood of acting on it, or of the event's occurring (so the mother

[1] Findings from factor analysis and imaging studies have led some researchers to classify hoarding as a related but distinct disorder; the latest edition of the DSM adopts this classification.

may take her thought of killing her child as providing evidence that she will act in that way, while another sufferer may take his thought of his mother's being hit by a car as making it probable that she will be hit by a car). Together, these dysfunctional beliefs or traits dispose sufferers to appraise a thought as very important, highly unacceptable, or as posing a threat for which the person is responsible, and therefore as something that must be suppressed or controlled (Wells 2000).

These appraisals explain the distress experienced: whereas one person might experience a passing thought of harming a loved one without undue worry, the patient is made highly anxious by the thought. It also explains the origin of compulsions. Compulsions are repetitive overt or mental actions that the person feels compelled to perform in response to the unacceptable thought. A compulsion reduces distress in one of several ways: by replacing the unacceptable thought with another, or by addressing their content (so, for instance, hand washing reduces contamination fears by providing reassurance that the person is not contaminated). However, the compulsions do not provide a successful response to obsessions over the longer run. They may become decreasingly effective at forestalling anxiety, leading to the need to repeat them, perhaps in ever more elaborated forms. One likely explanation for their decreasing effectiveness is that the compulsion may come to serve as a cue for the thoughts that it was initially designed to suppress.

There is an impressive body of evidence in support of the (meta)cognitive model of OCD (see Hezel & McNally in press for review). Not only are measures of these traits correlated with the severity of symptoms (with different dysfunctional beliefs correlated with specific symptoms), manipulating dysfunctional beliefs is effective both at increasing and decreasing severity, and cognitive therapy that addresses dysfunctional cognitions is an effective treatment (Fisher & Wells 2005; Wilhelm et al. 2009; Solem, Haland, Vogel et al. 2009). Just as impressively, Abromowitz, Khandker, Nelson et al. (2006) found that these traits in prospective parents (who are known to be especially vulnerable to developing OCD symptoms in the first months of parenthood) predicted intrusive thoughts about their infant three months after the birth (though other work has produced inconsistent results, with Coles & Horng 2006 replicating the results but Coles, Pietrefesa, Schofield & Cook 2008 failing to replicate them). But the (meta)cognitive model suffers from a grave problem. While there is strong evidence that a heightened sense of responsibility and/or thought-action fusion often co-occur with, and even plays a causal role in, OCD, the correlation is far from perfect. Many patients show little sign of abnormal appraisals (Taylor, Abramowitz, McKay, et al. 2006). That fact necessitates the search for a better explanation. On the account I will propose, the symptoms and the metacognitive dispositions typically found in OCD have a common cause, in dysfunctional allocation of attention as a result of excessively precise prior beliefs. I will argue that the obsessions (with harm and its prevention, with order and with cleanliness) and the rituals that many patients engage in to forestall the harms or the occurrence of the obsessions,

actually have the same cause as the appraisals which have been held by previous researchers to cause catastrophic misinterpretations of common thoughts.

2. Predictive Coding and OCD

A growing number of cognitive scientists and philosophers believe that the generation and updating of predictions is central to how cognition works (Friston 2010; Hohwy 2013; Clark 2016). At multiple levels of the processing hierarchy, from extremely low level mechanisms designed to 'expect' the firing of particular neurons or the receipt of particular neurotransmitters, all the way through to conscious thought, cognition involves the making of predictions about inputs and updating those predictions in the light of actual inputs. When a prediction is falsified (that is, when what is known as surprisal occurs) the organism seeks to explain away the error either by altering its expectations or – when the error is too large to be accommodated by relatively minor alterations in expectations – by passing it up the processing hierarchy. Errors from lower levels may thus serve as inputs at higher levels of cognition. Errors propagate up while predictions propagate down, constraining expectations and resolving prediction errors.

Models centred on predictions and updating in response to prediction errors have long been familiar in cognitive science. For instance, the best-known account of the Capgras delusion (the delusion that a familiar person has been replaced by a duplicate) involves prediction errors. On this account, roughly, the relevant mechanisms 'expect' a feeling of familiarity to accompany perception of a familiar person. Absence of that feeling (due to a lesion, say) constitutes a prediction error, and the delusional belief arises in an attempt to accommodate – explain away – the surprisal (this is standardly developed as a two-factor model of delusion, with the second factor consisting in a belief bias; as Hohwy 2013 has shown, there is no need for a second factor if we take the predictive coding framework seriously). The current wave of predictive coding centred models might be seen as generalizations of this account, now postulated to explain not only dysfunction and pathology but normal cognition as well (for instance, perception is held to occur when there is a sufficiently good match between expected input and actual input; Hohwy 2012, Clark 2016).

Predictive coding has proven a highly successful framework for explaining and illuminating psychopathologies like delusions (Hohwy 2013) and depersonalization (Seth, Suzuki & Critchley 2011). A prevalent view is that many psychiatric conditions can be thought of as reflecting an imbalance between the precision afforded to sensory information and the precision of prior beliefs. This view has found particular traction in autism and schizophrenia research; with a special focus on the failure to attenuate sensory precision – and compensatory increases in prior precision (see Adams et al. 2013 for

a full discussion). The view offered here has a nearer precedent in Edwards et al. 2012. These authors review the computational failures and false inference that may underlie functional medical symptoms (aka hysterical symptoms) using a predictive coding account that highlights aberrant precision control and implicit pathology in attending to various exteroceptive and interoceptive cues.

The power of these accounts encourages us to attempt to apply the predictive coding framework to other disorders which continue to resist explanation, like OCD. This is a challenge recently taken up by Moore (2015). Moore adapts Szechtman and Woody's (2004) account of OCD, which has it arising from a dysfunction of the *security motivational system* (SMS). The SMS is identified by Szechtman and Woody with a hardwired set of species-typical behaviors concerned with the protection of the self and of others; extending this account, Moore identifies the SMS with the set of cognitive mechanisms that fulfil this role. He suggests that the SMS works, in part, by generating multiple 'narratives' in response to a threatening stimulus, where these narratives concern how the threat might play out. These narratives facilitate adaptive response to threats, by being utilized in planning. Following Friston, Mattout & Kilner (2011), Moore suggests that the generation of narratives is not specific to the SMS, but a ubiquitous mechanism for predicting how the external world will unfold.

Moore suggests that organisms are likely to differ in the richness and creativity of the narratives they generate. It is adaptive to generate narratives, even those with a relatively low probability, if such narratives prepare the organism to respond to especially harmful threats. The dysfunction central to OCD, Moore suggests, consists in the generation of too many (objectively) low probability narratives, and too high an assignment of relative probability to these narratives.[2] Correlatively, the mechanisms constitutive of the SMS give too low a weight to more likely scenarios. In normal agents, and in typical situations in which events unfold as they should, inputs received after completion of a task are taken to confirm a particular narrative and a feeling of security (i.e., a resolution of the uncertainty that figures so prominently in OCD) – caused by the match between the prediction of the highest-probability narrative and the actual input – is generated. Even though the same observation is compatible with low probability narratives (the gas somehow turned on again after checking, say), they are ignored because

[2] Formally, when encoding a finite set of counterfactual narratives or hypotheses about what I will do next, the precision of these beliefs corresponds to their entropy. For example, beliefs with a low entropy or uncertainty assign a high probability to one alternative and a low probability to others. Conversely, if there are too many plausible prior beliefs about (catastrophic) narratives, then the probability distribution is more evenly distributed and the representation has a low precision or high entropy. Crucially, in biologically plausible models of action selection, precision plays a key role in action selection. In other words, increasing the precision selects one of many narratives that can then be subsequently used to prescribe behavior. In this context, there is ample evidence that modulatory neurotransmitters such as dopamine play a role in this action selection through an encoding of precision. We will return to this point later.

their probability is too low. In patients, the same observations are not taken to rule out low-probability narratives, because too high a probability is assigned to these narratives. As a result, patients may fail to experience the feeling of knowing or completion ('yedasentience'; Szechtman & Woody 2004). Yedasentience, here, is understood as confirmation of the narrative upon which the agent places the highest value; i.e., the narrative considered *a priori* to be most likely.

Moore situates this account within the familiar dual process framework, which divides cognitive processes into two systems or types (Stanovich 1999; Kahneman 2011). Type 1 processes are automatic, fast, undemanding of cognitive resources or attention and often unconscious; they are also relatively inflexible and stereotyped. Type 2 processes are effortfully initiated and sustained, slow, demanding of cognitive resources, and typically conscious. Type 2 processes are typically associated with the personal level of cognition; the level at which we ascribe mental properties (typically states, like beliefs and desires) to people, rather than mechanisms constitutive of them, whereas type 1 processes are more often associated with the subpersonal level at which information is processed without the person having access to its contents (Frankish 2009). The failure to generate a feeling of security proposed by Moore is subpersonal, but type 2 inference remains intact, and the person herself recognizes that the low probability narrative can safely be ignored. However, the SMS produces a signal that cannot easily be overriden, let alone ignored (presumably because it is tasked with a function that is evolutionarily central) and its signal is highly motivating. It generates motivations to (a) improve the precision of the predictions and (b) respond to the signaled threat. Most, though not all, OCD sufferers have good insight: Moore thinks he can explain this fact as well as failures of insight, when they occur. For the majority, system 2 inference predominates and the person has insight into their condition. But the urgent SMS signal causes belief-discordant behavior, and the mechanisms of cognitive dissonance reduction produce pressure to confabulate personal-level beliefs that would explain the behavior (note that cognitive dissonance itself seems amenable to a predictive coding-based explanation: the behavior in which the agent engages is discordant with their personal beliefs and generates a prediction error; belief update explains away the error).

Moore's account is speculative, but the insight that OCD might be explained within a predictive coding framework is compelling. Compulsions are (apparently) motivated by a conditional prediction: if a ritual is not performed, some disaster will befall the patient or a loved one. Handwashing is, apparently, aimed at avoiding catastrophic contamination that is otherwise predicted to occur; another ritual might be aimed at suppressing the thought that makes an immoral action or a feared event more likely. Hoarding might be aimed at preserving something in case it is urgently needed; the thought of disposing of the item causes anxiety due to the prediction that it will be needed. All of this might be perspicuously understood as involving the generation of narratives that encode predictions concerning how events might turn out, as Moore suggests. On the other hand (though Moore does not note the connection), the

gold standard in treatment of OCD, exposure and response prevention (ERP), might best be understood as aimed at correcting dysfunctional predictions. As the name suggests, ERP involves exposing a patient to a trigger of obsessions or compulsions and asking them to refrain from the behavior they use to dampen distress (so, for instance, someone suffering from contamination obsessions might be exposed to a feared stimulus, like a dirty washbasin, and then simply asked to refrain from washing for a defined period of time). ERP is very effective in bringing about lasting remission of symptoms in those patients able to engage effectively with it (the drop-out rate is high, due to its aversiveness for patients; Abramowitz, Taylor & McKay 2009). We can understand its efficacy within the predictive coding framework: ERP causes the relevant mechanisms to update their predictions of disastrous consequences when these consequences fail to follow from exposure.

While Moore's account is suggestive, it has a number of problems. Most significantly, understanding OCD as a disorder of the SMS predicts correlates of OCD – heightened anxiety and a general tendency toward overestimation of risks – which are indeed common in the disorder but by no means universal. I will suggest that Moore's central insight – that OCD involves giving narratives predictive of catastrophe excessive weight– is correct, but it does not arise from a hyperactive SMS. Rather the inappropriate assignment of probabilities arises from a prior dysfunction: dysfunctionally heightened and directed attention.

The key on my view lies in recognizing the parsimony of the predictive machinery utilized to understand the actions of self and others and to generate motor commands. The same models are utilized to infer from inputs to their likely causes (including the intentions and goals of other agents) and as inverse models that infer from desired goals to the motor commands required to realize them. These models are constructed by generating predictions: predictions about the expected movements of other agents and predictions about proprioceptive feedback from the agent herself (Friston et al. 2011). In effect, these models are pushmi-pullyu representations (Millikan 1995); representations of states and of ways of bringing them about at the same time. Representing a state of affairs (that I kill my baby, say) is representing the motor commands for bringing about that state: that is, the representation (in some sense genuinely) predicts my causing the state I fear.

OCD arises, I suggest, when heightened attention to actions or intrusive thoughts causes these pushmi-pullyu representations to become overly precise. There is a close connection between the optimization of precision in predictive coding and attention. In the perceptual domain, selecting precise prediction errors for subsequent (hierarchical) belief updating involves increasing the gain in proportion to precision – in exactly the same way that attentional selection and biased competition is thought to act (Desimone 1998; Feldman & Friston 2010). Similarly, in the context of selection among competing narratives or policies, it is necessary to increase the precision of prior beliefs about what one will do

next. This has been discussed in terms of dopaminergic function; both in the context of predictive coding and Markov decision processes (Hazy, Frank, & O'Reilly 2007; Friston et al 2013). Overly precise pushmi-pullyu representations will dispose the person both toward dysfunctional belief update and the prediction of unwanted actions and outcomes.

3. Excessive Precision of Representations in OCD

OCD arises, I suggest, when heightened attention to actions or to intrusive thoughts causes overly precise pushmi-pullyu representations. Intrusive thoughts automatically cause the generation of the motor commands needed for bringing them about: to represent oneself as striking one's child is to cause the generation of the associated motor representations (c.f. ideomotor theory; see Shin, Proctor & Capaldi 2010 for review), and even the representation of another person's bringing about a state of affairs involves generating some of the same motor commands. To represent oneself as performing a certain action is to be poised to perform that action. Famously, neurons in F5 respond both to perception of action and to the performing of actions (Gallese, Fadiga, Fogassi & Rizzolatti 1996). As Friston et al. (2011) emphasize, this is an expected consequence of the prediction minimization machinery, on which sensory and motor representations are deeply intertwined. In theoretical neurobiology, this generalization of the prediction minimization machinery to action is known as Active Inference.

The symptoms of OCD *and* the appraisal styles characteristic of patients both arise from the excessive precision of pushmi-pullyu representations, I suggest. First, the appraisals. If the same machinery represents action, the goals of action and the motor commands for the generation of action, then there is sense in which thought-action fusion is not a mistake: it is a veridical perception. To represent oneself (however fleetingly) as performing an action or as bringing about a state of affairs is to be poised to perform that action or realize that goal; to generate the motor representations required for action initiation. This might explain both predictive and moralized TAF. I *am* prepared to perform that action; that might explain why I see myself both as likely to bring it about and as the kind of person who is inclined to bring it about.

It is easy to see how an inflated sense of responsibility for outcomes might also arise from overly precise pushmi-pulllyu representations of catastrophic outcomes. Since the person represents herself as prepared to bring such outcomes about, she takes herself to be responsible for *whether* they come about. It is also easy to see how the need to control thoughts arises. Representing oneself as preparing to bring about a feared consequence is representing oneself as having dangerous thoughts. It is less obvious how perfectionism and intolerance for uncertainty might arise, but attention to the prediction error minimization machinery will allow us to understand it too. In doing so, we will begin to understand

how the overly precise pushmi-pullyu representations generates the symptoms of OCD as well as the appraisal style.

The generation of a prediction of impending disaster is the generation of large prediction error, relative to the model of the world that the agent values and which she aims to make actual or sustain. It is therefore highly motivating. It motivates behaviors aimed both at increasing the precision of the signal and at heading off the predicted disaster. The rituals and compulsions characteristic of OCD can very plausibly be conceptualized at aimed at both increasing precision and at prevention. Checking and rechecking is an attempt at rendering the signal more precise; ordering rituals may also be understood in this light (an orderly environment is a predictable environment).[3] Washing might be both an attempt at precisification – at simplifying the signal – and aimed at preempting disaster, while mental rituals might be aimed at replacing the thought that represents and prepares for bringing about that outcome. Now, the motivation to precisify automatically upregulates attention; in fact, attention is best thought of as a mechanism *for* precisification (Hohwy 2013). That is, the surprisal central to OCD orientates the person to the input and motivates them to seek as much detail as possible. They are highly motivated to seek certainty. This drive towards precise representations of narratives generates (or constitutes) an intolerance of uncertainty. This argument suggests that one central (possibly dopaminergic or neuromodulatory) deficit can be variously interpreted as a failure of optimal precisification, attentional selection, action selection – and their metacognitive consequences.

The cognitive and metacognitive dispositions characteristic of OCD therefore arise from the same machinery that generates symptoms, rather than causing them. As we have seen, however, there is some evidence that preexisting dispositions toward these traits predict later development of symptoms. While the evidence is inconsistent, it is plausible that a preexisting disposition toward these traits might be involved in setting the vicious circle of OCD going. A disposition toward intolerance of uncertainty or a heightened sense of responsibility might generate a need for extra care and therefore greater attention to the signal in the first place. Precisification fails to explain away the error; rather it heightens it: it ups the gain on a pushmi-pullyu representation which is normally unavailable. Hence precisification may set going a cycle of ever greater need to attend.

Why aren't all OCD patients subject to these cognitive and metacognitive dispositions? Insofar as the account advanced here is correct, and OCD centrally involves dysfunctionally heightened and directed attention to pushmi-pullyu representations, the root dysfunction is subpersonal, not personal. Having a subpersonal disposition to precisify beliefs about action will tend to cause or partly constitute personal

[3] Radomsky & Rachman (2004) found that non-clinical populations experienced less anxiety in an orderly environment.

level intolerance of uncertainty; having a sub-personal disposition to select catastrophic narratives in response to heighted attention to a pushmi-pullyu representation will tend to cause or to partially constitute a personal-level inflation of responsibility (and so on, for the other characteristic dispositions). But these subpersonal dispositions will not invariably bring about their personal-level correlates. The causal route from the first to the second might be disrupted, or countervailing pressures might prevent the first from constituting the second.

If representing a state of affairs involves the representation of the motor commands for bringing it about, but these thoughts occur in the non-clinical population as well as among patients, why does the OCD sufferer alone respond to them in a way that leads to the cycle of heightened attention and the compulsion to preempt or prevent action? Work on the neural correlates of OCD provides us with a clue here. fMRI work on the neural basis of OCD has produced data with a degree of concordance across studies that is among the highest for any psychiatric disorder (Pauls, Abramovitch, Rauch & Geller 2014). That work identifies a common circuitry as centrally implicated in OCD: a cortico-striato-thalamo-cortical circuit. Imaging studies (Menzies et al. 2008; Fitzgerald et al. 2011), connectivity studies (Harrison et al. 2009) and animal studies (Ahmari 2013; Burguiere, E., Monteiro, P., Feng, G. & Graybiel 2013) converge in identifying hyperactivity in this circuitry across OCD patients, regardless of symptomology.[4] Further support for this claim comes from treatment studies, with numerous studies showing decreased activation in this circuit following successful treatment, whether with cognitive behavioural therapy (Schwartz et al. 1996; Freyer et al. 2011) or psychopharmaceuticals (Saxena et al. 1999; Saxena et al. 2002).

This cortico-striato-thalamo-cortical circuity is widely held to be the same circuity which underlies conscious awareness, realizing the so-called global neuronal workspace (Dehaene & Naccache 2001; Dehaene, Changeux & Naccache 2011; see Levy 2014 for review). OCD, in short, seems to be a disorder of consciousness, centrally involving dysfunctionally heightened awareness of action. Summarizing the evidence, Milad and Rauch (2012) conclude that in OCD information pertaining to habitual actions that is processed in nonconscious cortico-striatal networks in healthy controls is instead processed in fronto-hippocampal circuitry; patients attend to aspects of their own actions which are normally carried out without awareness. As we have seen, attention is a mechanism for signal precisification: it allows for greater precision in the input. But heightened attention here is dysfunctional, because it brings with it excessive precision of pushmi-pullyu representations predictive of catastrophic outcomes. Precisification leads to a higher value being placed on the signal: because it

[4] Differences in the heritability of childhood onset OCD and adult onset OCD has led some researchers to think that they are distinct disorders (Pauls, Abramovitch, Rauch & Geller 2014). I take no stand on this issue, but whether the disorders are one or two, both patient groups exhibit this cortico-striatal hyperactivity.

is more precise, it is assigned a greater weight. This may set a vicious cycle of ever greater vigilance and ever heightened anxiety in motion. Note that there need be no suggestion that the patient is aware of the content of her pushmi-pullyu representations: she attends to her actions and to her thoughts thereby bringing it about that these representations are excessively precise.

It is worth attempting to set out the proposed causal mechanism for the generation of OCD and of the cognitive dispositions characteristic of OCD sufferers more explicitly.

(1) OCD begins with heightened attention to action and thoughts. What precipitates this heightened attention probably varies from person to person, but we can identify some likely contributors to it, since we know a great deal about attention. Attention is a mechanism for precisification of inputs; it is upregulated either by surprisal or by increasing the stakes such that errors are costlier. There is evidence that an increase in the stakes exacerbates OCD. Several studies have manipulated perceived responsibility for a task. For both clinical (Lopatka and Rachman 1995) and non-clinical (Ladouceur et al. 1995) samples, increased perceived responsibility increases anxiety and checking behaviors. Individual differences with regard to baseline responsibility can be expected to predispose to OCD; so, of course, might having responsibility given to one (recall the vulnerability of new parents to symptoms of OCD). Overestimation of threat can also be expected to upregulate attention, and indeed there is a correlation between (relative) overestimation of threat and OCD (Hezel & McNally in press). Anxiety might result from a heightened perception of responsibility or from an overestimation of threat (or from actual threats or trauma; it is worth noting that there is evidence of a link between OCD and trauma; Dykshoorn 2014), and anxiety too upregulates attention (Gerrans 2016 explains this upregulation with the framework of predictive coding).

(2) By whatever mechanism it comes to be upregulated, attention is dysfunctionally focused. It causes excessive precision of pushmi-pullyu representations. Since heightened attention is for increased precision, its failure to allay anxiety might itself constitute surprisal.

(3) Surprisal automatically engages attention, but since the problem arises from dysfunctionally enhanced attention, this merely exacerbates the problem: it causes greater weight to be placed on signals that predict disaster.

This proposed etiology helps to explain both the behaviors symptomatic of OCD *and* the dispositions characteristic of it. Many of the behaviors symptomatic of OCD can be explained as aimed at making signals more precise and therefore at minimization of surprisal. Checking, most obviously, is an attempt to sample the environment to increase precision in evidence, thereby allowing for model update or

action aimed at prediction error minimization. Other behaviors aim more directly at minimizing prediction errors by altering the environment rather than sampling it ("active inference"; Friston, Mattout & Kilner 2011). Ordering is an attempt to minimize environmental variation and thereby increase predictability, while cleaning rituals simultaneously precisify the incoming signal and remove contamination. Other behaviors are aimed more directly at suppressing or replacing the pushmi-pullyu representations which are over-attended: mental rituals, thought replacement and suppression aim at producing what the agent wants to see or feel rather than what he or she wants to do (Friston et al. 2010).

At the same time, some of the very same mental dispositions which predict vulnerability to OCD might themselves be caused by the dysfunctionally directed attention that lies at its heart. Attention to pushmi-pullyu representations automatically generates thought-action fusion, since such fusion is a veridical perception of these representations; pushmi-pullyu representations are simultaneously representations of states of affairs and directive of behavior. Attention to these representations also tends to generate a heightened sense of responsibility, since taking oneself to be poised to bring about a disastrous state of affairs may involve taking oneself to have a special responsibility for averting that state. If one's representations are poised to bring about disaster, further, the need to control them and an inflated sense of their importance follows directly. It is also easy to see how a drive to precisify sensory evidence against alternative (catastrophic) narratives can cause intolerance of uncertainty; high stakes decisions must be made in the light of the best – the least ambiguous – evidence. Indeed, there is evidence that repeated checking increases uncertainty about recalled information, which in turn increases a sense of responsibility for harm (Williams, Mugno, Franklin & Faber 2013).

It is also worth pointing out that the model seems to explain the relative efficacy of ERP. ERP deliberately causes the generation of pushmi-pullyu representations that carry information predictive of catastrophe. Having the patient refrain from suppressing that representation or taking action that preempts an apparently predicted event (e.g., washing his hands) causes the generation of surprisal relative to that prediction. This surprisal is passed on up the processing hierarchy in search of a model that explains it away. The result is model update; the adoption (or relearning) of a model on which the inputs observed do not predict catastrophe.[5] ERP is one treatment belonging to the broader family of

[5] Noggle (forthcoming) advances several arguments aimed at showing that ERP should not be understood as providing evidence that is disconfirmatory of a prediction. First, OCD sufferers already have ample disconfirmatory evidence available: for example, from repeated checking. Second, the timescale on which ERP works is incompatible with genuine disconfirmation. Sickness from having contaminated hands, for example, would not show up in the few hours of a single ERP session. Third, ERP works best when patients focus their attention on the feared stimulus, but disconfirmation does not require that we attend anywhere in particular. Finally, exposure in imagination has been found to be effective. While these considerations might constitute powerful objections to the claim that ERP works through personal level disconfirmation, the model here proposes subpersonal disconfirmation and

cognitive behavioral therapy (CBT), and there is good evidence for the efficacy of other applications of CBT for OCD (Fisher & Wells 2005; Fisher 2009). CBT for OCD centrally involves putting forward alternative explanations of observations to those entertained by patients, testing the validity and utility of appraisals and the necessity of rituals (Salkovskis 2007). These strategies may be understood as the provision of evidence in favor of a model that best explains the inputs, thereby minimizing surprisal.

As we saw, Moore (2015) adapted Szechtman and Woody's (2004) account, which understands OCD as arising from a dysfunctional SMS. Szechtman and Woody's account of OCD has come in for a great deal of criticism from other researchers. For instance, Taylor, McKay & Abramowitz (2005) argue that it can at best explain only a subset of the symptoms manifested by sufferers, and cannot explain why the dysfunction gives rise to compulsions rather than to generalized anxiety. Given the problems the model faces, it seems perilous to stake too much on its being vindicated. Note, however, that the proposal I have advanced here does not depend on the SMS hypothesis. The account is agnostic on the very existence of the SMS. Rather, it explains why patients give too great a weight to narratives predictive of catastrophe by way of overly precise pushmi-pullyu representations.

Edwards et al. (2012) propose an account of 'hysterical' (somatoform or psychogenic) symptoms that is closely related to the current proposal. On their view, these symptoms are produced by attentional processes. Patients allocate excessive attention to stochastically occurring sensations, either top-down or bottom-up, due to excessive precision in prior beliefs concerning these symptoms (this excessive precision might result from previous trauma, from somatic illness, even culturally mediated expectations). Because these sensations are attended and taken to confirm the prediction, the prior belief is reinforced and the problem is exacerbated. This proposal shares with the hypothesis put forward here the claim that symptomology is the product of the allocation of excessive attention to bodily sensations or motor representations that normally are unattended or only peripherally attended, where this allocation is caused by and subsequently reinforces excessive precision in prior beliefs. Edwards et al. note that their account predicts that patients will have an abnormal experience of agency, due to the allocation of excessive attention to the sensory consequences of movements; this is a prediction for which there is already evidence (Edwards et. al. 2011). They also predict that patients will perform better in the force-matching task (Shergill et al. 2003) that measures the attenuation of proprioceptive

update. The manifest contents of the fears of the patient need not match the subpersonal contents. The the predicted sickness, for instance, is unlikely to have the features that the agent herself would attribute to a genuine illness. It is unsurprising that heightened attention is conducive to successful treatment, since attention ups the gain on the signal and causes any resulting observations to be given proportionately greater weight in updating predictions. As we have seen, attention is dysfunctionally focused in OCD; treatment may involve redirecting it more adaptively. Finally, the fact that imaginal exposure works is also unsurprising: the mechanisms that generate the prediction may not have the resources to distinguish between veridical perception and other sources of inputs.

prediction errors. The account put forward here seems to make the same predictions: OCD patients allocate excessive attention to their actions and will therefore fail to attenuate prediction errors to the same degree as normal controls. There is independent evidence that forward models in OCD patients are dysfunctional in just the manner predicted, further supporting the account (Gentsch et al. 2012).

4. *Conclusion*

There is evidence suggesting that OCD is not one disorder but several. It is symptomatically heterogeneous, but individual patients show stability in symptom dimensions over time: contents may change, but symptom types tend to persist; furthermore, different symptom types seem to have distinctive biological correlates and inheritance patterns (Williams, Mugno, Franklin & Faber 2013). Despite this heterogeneity, I suggest that a common dysfunction gives rise to the disorder: dysfunctionally heightened and focused attention. Heightened attention to thoughts and actions causes overly precise pushmi-pullyu representations that, in turn, generate a prediction of catastrophe. This sets going a vicious cycle of increasing anxiety, heightening attention still further, and the performance of actions (overt and mental) to change the inputs or precisify the signals.

I do not take a stand on whether OCD is one disorder or many. There are many possible causal routes to dysfunctionally directed and heightened attention. Pre-existing anxiety, or other traits, or trauma, or perhaps dispositions better characterized subpersonally, might play a role in its etiology, with different traits and dispositions playing different roles in different classes of patients. Perhaps these facts give us grounds for distinguishing different kinds of disorders. Nevertheless, I suggest, they have a common core and a common cause. Given the capacity of the account I have sketched here to explain symptoms *and* the personal-level dispositions characteristic of sufferers, I believe it ought to be taken seriously as an explanation of the disorder.

Department of Philosophy
Macquarie University

References

Abramowitz, J. S., Khandker, M., Nelson, C. A., Deacon, B. J., & Rygwall, R. 2006. The role of cognitive factors in the pathogenesis of obsessions and compulsions: A prospective study. *Behaviour Research and Therapy* 44: 1361-1374.

Abramowitz, J.S., Taylor, S. & McKay, D. 2009. Obsessive-Compulsive Disorder. *The Lancet* 8; 374 (9688): 491-9.

Adams, R.A., Stephan, K.E., Brown, H.R., Frith, C.D. & Friston, K.J. 2013. The computational anatomy of psychosis. *Frontiers in Psychiatry* 4: 47.

Ahmari, S. E. et al. 2013. Repeated cortico-striatal stimulation generates persistent OCD-like behavior. *Science* 340: 1234–1239.

Burguiere, E., Monteiro, P., Feng, G. & Graybiel, A. M. 2013. Optogenetic stimulation of lateral orbitofronto-striatal pathway suppresses compulsive behaviors. *Science* 340: 1243–1246.

Clark, A. 2016. *Surfing Uncertainty*: *Prediction, Action, and the Embodied Mind.* Oxford: Oxford University Press.

Coles, M. & Horng, B. 2006. A Prospective Test of Cognitive Vulnerability to Obsessive-compulsive Disorder. *Cognitive Therapy and Research* 30: 723-734.

Coles, M.E., Pietrefesa, A.P., Schofield, C.A., & Cook, L.M. 2008. Predicting changes in OC symptoms over time: A prospective test of cognitive models of OCD. *Cognitive Therapy and Research* 32: 657-675.

Corlett, P.R., Frith, C.D. & Fletcher, P.C. 2009. From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology* 206: 515-30.

Dehaene, D. & Naccache, L. 2001. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79: 1-37

Dehaene, S., Changeux J.-P, and Naccache L. 2011. The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architectures to Clinical Applications. In S. Dehaene and Y. Christen (eds.), *Characterizing Consciousness: From Cognition to the Clinic?* Berlin: Springer-Verlag, 55-84.

Desimone, R. 1998. Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London B, Biological Sciences* 353: 1245-1255.

Dykshoorn K.L. 2014. Trauma-related obsessive-compulsive disorder: a review. *Health Psychology and Behavioral Medicine* 2: 517-528.

Edwards, M.J., Adams, R.A., Brown, H., Pareés, I. & Friston, K. 2012. A Bayesian account of 'hysteria'. *Brain* 135: 3495-3512.

Edwards M.J., Moretto G., Schwingenschuh P., Katschnig P., Bhatia K.P. & Haggard P. 2011. Abnormal sense of intention preceding voluntary movement in patients with psychogenic tremor. *Neuropsychologia* 49: 2791–3.

Feldman H. & Friston K.J. 2010. Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience* 2 :215.

Fisher, P.L. & Wells, A. 2005. How effective are cognitive and behavioral treatments for obsessive–compulsive disorder? a clinical significance analysis. *Behaviour Research and Therapy* 43: 1543–1558.

Fisher, P.L. 2009 Obsessive Compulsive Disorder: A Comparison of CBT and the Metacognitive Approach. *International Journal of Cognitive Therapy* 2: 107-122.

Fitzgerald, K. D. et al. 2011. Developmental alterations of frontal-striatal-thalamic connectivity in obsessive-compulsive disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* 50: 938–948.e3.

Frankish, K. 2009. Systems and levels: Dual-system theories and the personal—subpersonal distinction. In Jonathan Evans and Keith Frankish (eds), *In Two Minds: Dual Processes and Beyond.* Oxford: Oxford University Press, pp. 89–107.

Freyer, T. et al. 2011. Frontostriatal activation in patients with obsessive-compulsive disorder before and after cognitive behavioral therapy. *Psychological Medicine* 41: 207–216.

Friston, K. 2010. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience* 11: 127-138.

Friston K.J., Daunizeau J., Kilner J. & Kiebel S.J. 2010. Action and behavior: a free-energy formulation. *Biological Cybernetics* 102: 227–60.

Friston, K, Mattout, J., &Kilner, J. 2011. Action understanding and active inference. *Biological cybernetics* 104: 137-160.

Friston K., Schwartenbeck P., FitzGerald T., Moutoussis M., Behrens T., Dolan RJ. 2013. The anatomy of choice: active inference and agency. *Frontiers in Human Neuroscience* 7:98.

Gallese V, Fadiga L, Fogassi L, Rizzolatti G. 1996. Action recognition in the premotor cortex. *Brain* 119: 593-609.

Gentsch, A., Schütz-Bosbach, S., Endrass, T. & Kathmann, N. 2012. Dysfunctional Forward Model Mechanisms and Aberrant Sense of Agency in Obsessive-Compulsive Disorder. *Biological Psychiatry* 71: 652-659.

Gerrans, P. 2016. All the self we need. In Thomas Metzinger and Jennifer Windt (eds), *Open MIND: Philosophy and the Mind Sciences in the 21st Century*. Cambridge, Mass.: The MIT Press.

Gibbs N. 1996. Nonclinical populations in research on obsessive-compulsive disorder. *Clinical Psychology Review* 16: 729–73.

Haker, H., Schneebeli, M. & Stephan, K.E. 2016. Can Bayesian Theories of Autism Spectrum Disorder Help Improve Clinical Practice? *Frontiers in Psychiatry* 7: 107.

Harrison, B. J. et al. 2009. Altered corticostriatal functional connectivity in obsessive-compulsive disorder. *Archives of General Psychiatry* 66: 1189–1200.

Hazy, T.E., Frank, M.J. & O'Reilly, R.C. 2007. Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philosophical Transactions of the Royal Society of London B, Biological Sciences* 362: 1601-1613.

Hezel D.M. & McNally R.J. In Press. A Theoretical review of cognitive biases and deficits in obsessive-compulsive disorder. *Biological Psychology.* doi:10.1016/j.biopsycho.2015.10.012

Hohwy, J. 2012. Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology* 3: 96

Hohwy, J. 2013. *The Predictive Mind.* Oxford: Oxford University Press.

Kahneman, D. 2011. *Thinking Fast and Slow*. London: Allen Lane.

Ladouceur, R., Rheaume, J., Freeston, M.H., Aublet, F., Jean, K., Lachance, S., Langlois, F., & de Pokomandy-Morin, K. 1995. Experimental manipulations of responsibility: an analogue test for models of obsessive-compulsive disorder. *Behaviour Research and Therapy*, 33: 937-946.

Levy, N. 2014. *Consciousness and Moral Responsibility.* Oxford: Oxford University Press.

Lopatka, C., & Rachman, S. J. 1995. Perceived responsibility and compulsive checking: And experimental analysis. *Behaviour Research and Therapy* 33: 673–684.

Menzies, L. et al. 2008. Integrating evidence from neuroimaging and neuropsychological studies of obsessive-compulsive disorder: the orbitofrontostriatal model revisited. *Neuroscience and Biobehavioral Reviews* 32: 525–549.

Milad, M.R & Ruach, S.L. 2012. Obsessive-compulsive disorder: beyond segregated cortico-striatal pathways. *Trends in Cognitive Science* 16: 43-51

Millikan, R.G. 1995. Pushmi-Pullyu Representations. *Philosophical Perspectives* 9: 185-200.

Moore, P.J. 2015. A predictive coding account of OCD. *arXiv* preprint. arXiv:1504.06732.

Robert Noggle. Forthcoming. Belief, quasi-belief, and obsessive-compulsive disorder. *Philosophical Psychology*.

Obsessive-Compulsive Cognitions Working Group. 2003. Psychometric validation of the Obsessive Belief Questionnaire and Interpretation of Intrusions Inventory-Part 1: Factor analyses and testing of a brief version *Behaviour Research and Therapy* 43: 863-878.

Obsessive-Compulsive Cognitions Working Group. 2005. Psychometric validation of the Obsessive Belief Questionnaire and Interpretation of Intrusions Inventory-Part 2: Factor analyses and testing of a brief version *Behaviour Research and Therapy* 43: 1527–1542.

Pauls, D.L., Abramovitch, A., Rauch, S.L., Geller, D.A. 2014. The neuroscience of obsessive-compulsive disorder: An integrative genetic and neurobiological perspective. *Nature Reviews Neuroscience* 15: 410-424.

Radomsky A. & Rachman, S. 2004. Symmetry, ordering and arranging compulsive behavior. *Behaviour Research and Therapy* 42: 893–913.

Salkovskis P.M. 1985. Obsessional-compulsive problems: A cognitive-behavioural analysis. *Behaviour Research and Therapy* 23: 571–83.

Salkovskis P.M. 2007. Psychological treatment of obsessive–compulsive disorder. *Psychiatry* 6: 229-233.

Saxena, S. et al. 1999. Localized orbitofrontal and subcortical metabolic changes and predictors of response to paroxetine treatment in obsessivecompulsive disorder. *Neuropsychopharmacology* 21: 683–693.

Saxena, S. et al. 2002. Differential cerebral metabolic changes with paroxetine treatment of obsessivecompulsive disorder versus major depression. *Archives of General Psychiatry* 59: 250–261.

Schwartz, J. M. et al. 1996. Systematic changes in cerebral glucose metabolic rate after successful behavior modification treatment of obsessive-compulsive disorder. *Archives of General Psychiatry* 53: 109–113.

Seth, A., Suzuki, K. & Critchley, H.D. 2011. An Interoceptive Predictive Coding Model of Conscious Presence. *Frontiers in Psychology* 2: 395.

Shergill S.S., Bays P.M., Frith C.D. & Wolpert D.M. 2003. Two eyes for an eye: The neuroscience of force escalation. *Science* 301: 187.

Shin, Y.K., Proctor, R.W. & Capaldi, E. J. 2010. A review of contemporary ideomotor theory. *Psychological Bulletin*136: 943-974.

Solem, S., Haaland, A.T., Vogel, P.A.; Hansen, B. & Wells, A. 2009. Change in metacognitions predicts outcome in obsessive-compulsive disorder patients undergoing treatment with exposure and response prevention. *Behaviour Research and Therapy* 47: 301-307.

Stanovich, K. 1999. *Who Is Rational? Studies of Individual Differences in Reasoning*. Mahwah: Lawrence Erlbaum Associates.

Szechtman, H. & Woody, E. 2004. Obsessive-compulsive disorder as a disturbance of security motivation. *Psychological Review* 111: 111-127.

Taylor S., McKay D., Abramowitz, J.S. 2005. Is obsessive-compulsive disorder a disturbance of security motivation? Comment on Szechtman and Woody (2004). *Psychological Review* 112: 650-7.

Taylor S., Abramowitz J.S., McKay D., et al. 2006. Do dysfunctional beliefs play a role in all types of obsessive-compulsive disorder? *Journal of Anxiety Disorders* 20: 85–9.

Veale, D. & Roberts, A. 2014. Obsessive-compulsive disorder. *British Medical Journal* 348: g2183.

Wells, A. 2000. *Emotional disorders and metacognition: Innovative cognitive therapy*. Chichester, UK: Wiley.

Wilhelm, S., Steketee, G., Fama, J.M., Buhlmann, U., Teachman, B.A. & Golan, E. 2009. Modular Cognitive Therapy for Obsessive-Compulsive Disorder: A Wait-List Controlled Trial. *Journal of Cognitive Psychotherapy* 23: 294–305.

Williams, M.T., Mugno, B., Franklin, M. & Faber, S. 2013. Symptom Dimensions in Obsessive-Compulsive Disorder: Phenomenology and Treatment Outcomes with Exposure and Ritual Prevention. *Psychopathology* 46: 365–376.