Running head: BETTER SAFE THAN SORRY

# Better safe than sorry:

# A common signature of general vulnerability for psychopathology

Omer Van den Bergh*

Health Psychology, University of Leuven, Leuven, Belgium

Jos Brosschot

Health, Medical and Neuropsychology Unit, Institute of Psychology, Leiden University, Leiden, The Netherlands

Hugo Critchley

Division of Psychiatry, Brighton and Sussex Medical School, University of Sussex, Brighton, UK

Julian F. Thayer

Department of Psychological Science, University of California, Irvine, USA

Cristina Ottaviani

Department of Psychology, Sapienza University of Rome, Rome, Italy

Neuroimaging Laboratory, Santa Lucia Foundation, Rome, Italy

*Correspondence concerning this article should be addressed to Prof. Omer Van den Bergh, Health Psychology, Tiensestraat 102, University of Leuven, Belgium. Contact: omer.vandenbergh@kuleuven.be

**Accepted for publication in**

**Perspectives in Psychological Science**

**22-Apr-20**

**Abstract**

Several labels such as neuroticism, negative emotionality and dispositional negativity indicate a broad dimension of psychopathology. However, largely separate research lines have developed, often disorder-specific, that focus on different cognitive and affective characteristics that are associated with this dimension, such as perseverative cognition (worry, rumination), reduced autobiographical memory specificity, compromised fear learning, and enhanced somatic symptom reporting. In this paper, we present a theoretical perspective within a predictive processing framework in which we trace these phenotypically different characteristics back to a common underlying "better safe than sorry"- processing strategy. This implies information processing that tends to be low on sensory-perceptual detail, giving room to threat-related categorical priors to dominate conscious experience and to chronic uncertainty/surprise due to a stagnated error reduction process. This common information processing strategy has beneficial effects on the short term, but important costs on the longer term. Viewed from this perspective, we suggest that the phenomenally distinct cognitive and affective psychopathological characteristics mentioned above represent the same basic processing heuristic of the brain, and are only different as regards the particular type of information involved (e.g. information in working memory, in autobiographical memory, in the external and internal world). Clinical implications of this view are discussed.

Keywords: Neuroticism, negative emotionality/affectivity, dispositional negativity, fear learning, autobiographical memory, somatization, perseverative cognition

One of the most challenging problems for clinicians and researchers of psychopathology is to capture the large variability in symptom profiles into a parsimonious, comprehensive and clinically useful classification system. Historically, categorical descriptive schemas such as the Diagnostic and Statistical Manual of Mental Disorders (DSM, American Psychiatric Association) and the International Statistical Classification of Diseases and Related Health Problems (ICD, World Health Organization) have dominated this endeavor. However, categorical models of psychopathology and psychiatric disorders are increasingly criticized for several reasons, such as the use of consensus- rather than evidence-based categories, arbitrary thresholds to delineate a categorical diagnosis, the large heterogeneity of symptoms and processes within categories, their limited reliability and substantial comorbidity (Kotov et al., 2017). On the other hand, a quantitative nosology based on empirical data analysis is rapidly developing, showing that the observed variation in psychopathology can be parsimoniously modeled using a limited number of hierarchically arranged dimensions (Caspi et al., 2014; Kotov et al., 2017). For example, analyzing the correlational structure among 11 first-order dimensions of psychopathology, Lahey, Krueger, Rathouz, Waldman and Zald (2017, p. 161) suggested one psychobiological dimension that increases the risk for psychopathology and stated that “negative emotionality lies at the heart of the general factor of psychopathology”. Investigations using a broader database suggested a large p-factor (Caspi et al., 2014; Caspi & Moffitt, 2018) as a dimension indicating the liability for any mental disorder, as well as for comorbidity, persistence and symptom severity with disordered thought at the extreme end. Recently, a dimensional hierarchical taxonomy of psychopathology was proposed to guide mental health research (HiTOP; Conway et al., 2019) describing a general psychopathology factor at the top and large spectra at the level below it reflecting dimensions of individual differences as identified in personality research. These spectra in turn aggregate syndromes and disorders and - further down the hierarchy - more narrow signs, symptoms and components of psychopathological functioning to allow different degrees of specificity when describing the pathological state of an individual.

In the present paper, we will focus on the internalization spectrum as described in the HiTOP structure, which in personality research corresponds with different terms such as neuroticism (Eysenck, 1947; Zinbarg, Mineka, Bobova, Craske, Vrshek-Schallhorn, Griffith et al., 2016), negative emotionality (Eisenberg et al., 2005), trait negative affectivity (Watson & Clark, 1984), and dispositional negativity (Shackman et al., 2016a,b) (abbreviated as N/NE in the remainder of this paper). There are several good reasons to focus on this trait. First, N/NE is associated with less favorable conditions and outcomes in almost every domain of life, including education, work, relational stability, and various indicators of physical health, resulting in substantial individual and societal burden and excessive costs (Cuypers et al., 2010; Shackman et al., 2016a). Second, N/NE is also critically involved in causing and maintaining psychopathology: N/NE is not only associated with a large array of disorders such as anxiety, mood and substance abuse disorders, with the amount of comorbidity and with a less favorable prognosis, it is also a powerful predictor for the development of these disorders in longitudinal studies (Shackman et al., 2016a; Hur et al., 2019). Third, recent meta-analytic evidence has shown that N/NE can be altered with different types of therapy with similar moderate effect sizes (e.g. Cohen's d=.50 -.60 for cognitive-behavioral, supportive or mixed therapies; Roberts et al., 2017). This is interesting because targeting a single transdiagnostic dimension such as N/NE is more efficient than developing separate treatments for a panoply of co-morbid categorical disorders. At the same time, it suggests that a clear understanding of the processes involved in this transdiagnostic trait is of major importance in order to make progress in preventing its consequences and improving treatment effects.

The goal of the present paper is to use a novel theoretical perspective on how the brain processes information to suggest a processing heuristic that is at the core of N/NE. We call this heuristic a "better safe than sorry" (BSTS) processing strategy. The label BSTS is not new in this context: earlier it has been used occasionally to indicate the rationale behind some N/NE-related mechanisms such as attentional and interpretational bias. However, it has not been understood as the central organizing principle at the core of N/NE, as is shown by the absence of this concept in

several important conceptual papers on N/NE (Barlow et al., 2014a; Barlow et al., 2014b; Shackman et al., 2016a,b; Hur et al., 2019). By taking a predictive processing perspective, we want to show that BSTS causes a stagnated error reduction process that – depending on the specific type of information involved - results in different phenomenal characteristics of N/NE that up to now have been investigated in largely separate research lines. By putting BTST at the core of N/NE and by demonstrating how separate phenomena can be understood as representing the same processing heuristic of the brain, we aim at contributing to a deeper and more parsimonious understanding of N/NE. We also want to show how this way of understanding may improve treatment by suggesting which treatment components are critical and why.

Before elaborating on this predictive processing perspective and its potential implications to understand N/NE, we will briefly summarize current views on this trait.

## 1. Conceptualizations of N/NE as a dimension of psychopathology

### *1.1. The nature of N/NE*

N/NE appears as a broad and stable disposition, resulting from both genetic and environmental factors and their interaction, to appraise situations as more threatening, to hold negative anticipations about oneself and the world, to experience negative mood states and emotions, and to show poor emotion regulation (Watson & Clark, 1984; McCrae & Costa, 2003; Ormel et al., 2012, 2013; Barlow, Ellard, Sauer-Zavala, Bullis, & Carl, 2014b). N/NE is different from a negative affective state in response to environmental factors. In general, people high on N/NE show more affective instability, that is, they show more moment-to-moment affective fluctuations, even between feelings of different quality (Kuppens, Van Mechelen, Nezlek, Dossche & Timmermans, 2007). Within a constructionist view on emotion (Barrett, 2017), N/NE should be seen as a valenced background state reflecting a rather chronic imbalance between the necessary and actual /anticipated resources to secure basic needs (e.g. growth, survival and reproduction) rather than as a categorical constructed emotion.

N/NE is associated with more vigilance towards potentially negative information, behavioral inhibition and greater intolerance for uncertain, ambiguous and uncontrollable situations, and a lower threshold to react with avoidance and escape responses to threat and negative information, including one's own negative emotions (Barlow et al., 2014a; Gray & McNaughton, 1996). N/NE is also associated with an array of behavioral characteristics such as attentional (Shackman et al., 2016a,b) and interpretational bias (Mathews, Ridgeway, Cook & Yiend, 2007), worrying, ruminating and catastrophizing (Ehring & Watkins, 2008), enhanced symptom reporting (Van den Bergh, Witthöft, Petersen, & Brown, 2017), reduced autobiographical memory specificity (Walker, Yancu & Skowronski, 2014) and compromised fear learning mechanisms, such as poor safety learning, enhanced generalization of threat perception, and poor extinction learning (Gazendam, Kamphuis & Kindt, 2013; Haaker, et al., 2015). It is likely that several of these characteristics rely in part on deficits in executive function and cognitive control (Hur et al., 2019).

In recent years, both animal and human research has shown that N/NE is associated with altered neural structure and function (Depue, 2009; Hariri, 2009; Ormel et al., 2013). Recent comprehensive reviews by Shackman and colleagues (Shackman et al., 2016 a,b; Hur et al., 2019) document elevated responding to threat in several brain areas, such as the amygdala, hippocampus, insula, bed nucleus of the stria terminalis (BNST), mid-cingulate and orbitofrontal cortex, and periaqueductal gray. Consistent evidence further suggests that the amygdala is a coordinating brain area involved in both resting state and threat-induced responses that are associated with elevated N/NE. On the one hand, the amygdala receives input from sensory (thalamic), contextual (hypocampal) and evaluative/regulatory (prefrontal/insular) structures, which flows from the ventral areas of the amygdala to the central (Ce) and dorsal-posterior parts and to the bed nucleus of the stria terminalis (BST). On the other hand, the latter structures (Ce, BST) coordinate the behavioral (e.g. avoidance, inhibition), physiological (e.g. autonomic, neuroendocrine) and cognitive (e.g. hypervigilance, attentional bias) responses characterizing negative affective states.

Interestingly, acute stressors alter amygdala functional connectivity and potentiate amygdala responses to threat (Cousijn et al., 2010; Hermans et al., 2017) and these effects are stronger in persons with N/NE (Everaerd et al., 2015) suggesting sensitization of the amygdala. Repeated acute and/or chronic stress may also lead to neural changes in other brain areas, including the hippocampus and prefrontal cortex (McEwen & Gianaros, 2011; McEwen, Nasca & Gray, 2016). These might impact larger brain networks subserving executive function and cognitive control such as the frontoparietal (dlPFC, intraparietal sulcus) and cingulo-opercular network (midcingulate cortex, anterior insula, frontal operculum) (Li et al., 2017). Emerging evidence suggests that frontoparietal areas such as dlPFC are less efficient in allocating resources to execute function tasks in persons with N/NE (see Hur et al., 2019, for a review).

*1.2. N/NE and psychopathology*

A major scientific challenge remains, however, to parsimoniously conceptualize N/NE as a general transdiagnostic psychobiological dimension in order to solve the multifinality problem (how a transdiagnostic factor causes different disorders) and the problem of divergent trajectories (why different persons with the same transdiagnostic factor develop different disorders) (Nolen-Hoeksema & Watkins, 2011). The latter authors suggested to solve these problems by sorting a large number of separate processes into a coherent framework consisting of distal, proximal and moderating (risk) factors (Nolen-Hoeksema & Watkins, 2011). In another attempt, Barlow and coworkers (Barlow et al., 2014b) describe a complex of vulnerability factors (triple vulnerability theory) involving, first, a biological (heritable) vulnerability factor predisposing towards hyperexcitability in emotion-related brain structures; second, a (learned) psychological vulnerability factor representing “a pervasive sense of unpredictability and uncontrollability in relation to life events and a perceived inability to cope with negative outcomes from such life events” (p.484), and, third, a (learned) factor that determines why an individual becomes concerned about a particular type of threat and, thus, develops a particular type of disorder. Also the Research Domain Criteria (RDoC) try to identify and integrate dimensional constructs at multiple levels of measurement (from

genes to self-reports) in different domains of functioning (Cuthbert & Insel, 2013). For example, Lang and coworkers suggested a dimension of psychophysiologic responding in the negative valence system that cuts across DSM-based clinical anxiety diagnoses representing diminishing defensive reactivity (or blunting of the defensive system) with decreasing focal fear and increasing negative affectivity, functional interference and distress of the patients (McTeague & Lang, 2012; Lang, McTeague & Bradley, 2016).

These approaches have created coherent conceptual schema's elucidating how general threat sensitivity may lead to various psychopathological phenomena through the operation of an array of empirically established processes. However, in this paper we suggest that taking a new and more radical functional perspective may further deepen our understanding in a more parsimonious way. In an evolutionary perspective, N/NE as a trait can be seen as an adaptive response to repeated experiences with threat involving a recalibration of the threshold for threat detection and concomitant vigilance and hyperarousal. An optimal threshold for threat detection depends on two factors: the probability of an aversive event and the relative payoffs of the four possible outcomes of detection (i.e. true and false positives and negatives) (Nettle & Bateson, 2012). Given the typically much larger costs of false negatives (missing threat) compared to false positives (metabolic expenditure), it is adaptive to recalibrate thresholds to lower levels in conditions where aversive experiences are likely, representing a "better safe than sorry"- strategy to cope with threat. Because of the autocorrelated nature of the person-environment system within person and across generations, it is no wonder that recalibration to lower thresholds has been observed in animals and humans, especially after repeated (early) adverse experiences, and that it can be transferred to next generations (Nettle & Bateson, 2012; Hanson, Hariri, & Williamson, 2015; McEwen et al., 2016; Baker, Cesa, Gatz, & Mellins, 1992; Hariri & Holmes, 2006, Kendler, Prescott, Myers, & Neale, 2003). Recalibration occurs in both animals and humans, suggesting a pervasive change in the way living animals construe and behave in the world that is not always captured by assessing cognitive operations and/or verbal expressions. In addition, it suggests that bias should not be considered

irrational and dysfunctional, but rather quite rational and adaptive in view of a person's history (Gilbert, 1998; Nemeroff, 2013).

In line with this functional perspective, a recent theory (Generalized Unsafety Theory of Stress, GUTS; Brosschot, Verkuil, & Thayer, 2016, 2017, 2018) suggested that the default state of organisms is to expect and prepare for threat, and that only the perception of safety can inhibit this default state. High N/NE individuals would be unable to perceive safety resulting in a chronic state of threat, even when none is present. The authors consider different behavioral characteristics such as attentional bias, negative interpretation biases, perseverative thinking, poor safety learning and chronic stress to represent a strategy of "erring on the safe side". However, whereas GUTS describes why these phenomena should be interpreted so, it does not explain how information is processed representing this strategy and how precisely this leads to the behavioral phenomena that can be observed in clinical practice.

Clarifying this "how" in terms of processing heuristics in the brain is the purpose of the present paper. We suggest that interpreting N/NE and associated cognitive, affective and behavioral phenomena within a predictive processing perspective elucidates the centrality and communality of a "better safe than sorry" (BSTS) strategy in a way that would not be apparent without these new conceptualizations. Predictive processing accounts entail a radical constructivism in which the brain is seen as an organ that actively constructs a model of reality from noisy input using information it already has. Prediction signals from models in the brain are matched with sensory input resulting in prediction errors that are fed back to improve the adaptivity of these models when making perceptual inferences and actively navigating in the environment. In a broad sense, we suggest that high threat sensitive persons tend to abort this error reduction process in a premature phase. While this strategy may be rewarded by beneficial effects on the short term – it helps to quickly classify information about threat and negative valence – it leads to insufficient updating of their model of reality resulting in persistent deviations between expected and actual input. We suggest that such a "stagnated error reduction process" is the core of the psychobiological dimension of N/NE and that

phenomenally different clinical characteristics that are associated with it are at a more fundamental level implementations of one and the same process.

Before elaborating on this, a brief introduction to predictive processing models will be given (for more elaborate introductions, see Friston, 2010, 2013; Friston et al., 2017; Hohwy, 2012, 2013; Clark 2013; Seth, 2013; Barrett & Simmons, 2015; Stephan, Manjaly, Mathys, Weber, Paliwal, Gard et al., 2016; Wiese & Metzinger, 2017; Van den Bergh, Witthöft, Petersen, & Brown, 2017).

## 2. A prediction processing model

Rooted in theories of inference and control in biological systems, new conceptualizations of the brain have emerged in computational neuroscience, the implications of which are currently being explored in a growing number of area's (Stephan et al. 2016; Barrett, 2017). Conceiving of the brain as "an ever-active hierarchical prediction machine" striving to minimize prediction error has important implications for a functional interpretation of perception, attention, emotion, thought, language and action and more widely for our understanding of mind, experience and agency, leading several authors to qualify this perspective as a paradigm shift (Hohwy, 2013; Clark, 2013; Lupyan & Clark, 2015; Barrett, 2017).

### *2.1. Predictive processing, perception and action*

According to the predictive processing framework, a basic task of the brain is to construct an adaptive model of the (external and internal) world, while its only source of information to do so is the spatial and temporal patterning of its own neural activity. Models are considered adaptive when they allow to quickly and efficiently infer the sources of stimulation (perception) as well as to predict future states and consequences of actions (planning and action). In order to achieve this goal, the brain uses information from neural activity that is triggered by peripheral input (sense organs and receptors in the peripheral body), but also from neural activity that is generated by the brain itself, reflecting previous experiences and "built in" information that act as predictions. The theory of

predictive coding or processing specifies how the brain makes sense of the stimulation it receives given certain expectations, using principles akin to Bayesian inference.

Input from the (external or internal) world leads to two counterflowing streams of neural activation across several hierarchical levels of the brain: stimulation by peripheral input (called "likelihood") interacts with activations generated by the brain ("generative model") that act as probabilistic predictions of the input ("prior beliefs") within a specific context, that is, estimate a likelihood of what the new input will be given previous experience. It should be noted though that prior beliefs are implicit assumptions of the brain that in most cases are not accessible to consciousness. So, they can be quite different from conscious beliefs. The discrepancy between predicted and actual inputs across multiple hierarchical levels — from low-level sensory input to high-level abstractions — results in prediction errors that are propagated throughout the system in a process of error minimization. Eventually, the brain will settle on/infer a posterior model that represents the most likely model of the stimulation. For example, if one is waiting for Jeff in a crowded street, the brain generates neural patterns acting as prior beliefs that will facilitate spotting Jeff in the crowd (example from Pezzulo, Maisto, Barca, & Van den Bergh, 2019, p.3).

The system realizes error minimization in three ways: (a) by adapting the prior beliefs to accommodate the actual input (belief update); (b) by actively operating on the world and generating input that fits the prior beliefs (so-called "active inference"); and (c) by changing how the brain samples (or attends to) sensory input (Barrett & Simmons, 2015). Active inference acknowledges that perception does not passively wait for sensory input but depends on action (e.g. active sampling) to produce input. For example, waiting for Jeff may prompt the person to move towards a location providing a better overview of the passing crowd and/or to increase the scanning rate generating more detailed information to help spotting him (Pezzulo, Maisto, Barca & Van den Bergh, 2019). In a broader sense, however, it refers to any kind of perceptual, somatovisceral and behavioral (goal-directed) responses to produce input that is consistent with the expectations specified in the generative model in order to minimize prediction error. Active inference therefore involves "policy

selection" to reach epistemic and/or pragmatic goals (Friston, FitzGerald, Rigoli, Schwartenbeck & Pezzulo, 2016). Eventually, this process of error minimization settles on posterior beliefs that best account for the prediction errors and that reflects the combined influence of priors and the actual input. Conscious experience is thought to correspond to the posterior beliefs that are the most likely explanation of what is happening in the world (Hohwy, 2012).

An interesting consequence of this perspective is that it integrates perception, action and physiological regulation within the same theoretical account (see Smith, Thayer, Khalsa, & Lane, 2017, for a hierarchical predictive coding model describing how multiple neurovisceral interactions regulate heart rate variability in interaction with physiological and psychological demands).

2.2. Precision and precision control

Priors, prediction errors, and posterior beliefs are conceived of as probability distributions that represent statistical regularities in neural activity with a mean and a variance. The inverse of the variance of these distributions is its precision. Highly precise priors and prediction errors reflect that a neural pattern has a high probability of being associated with a particular input, and conversely for low precise priors and prediction errors. For example, if Jeff is unusually tall, both priors and prediction errors representing Jeff's height are highly precise, resulting in a quick and reliable recognition of Jeff. The relative impact of the distributions representing prior beliefs (i.e., predictions) versus the distributions representing the input (i.e., the likelihood of what is present) on the posterior beliefs (i.e., what the system concludes is present) will be determined by the precision of the distributions. For example, when it is dark, there is a high probability to recognize Jeff in any tall person, reflecting a strong effect of the prior on the eventual perception. Conversely, on a sunny day it is less likely to take any tall person for Jeff and this likelihood is even further reduced if one is not waiting for Jeff.

Interestingly, because the brain cannot know whether any residual prediction error represents random information or is amenable to further minimization, it has to learn the conditions under which particular models are likely to be adaptive. This is accomplished by developing context-

dependent expectations about the precision of its inputs, which in turn determine how much weight is given to the prediction errors in the process ("precision optimization"; Hohwy, 2012). The implication is that contextual cues may have an important impact on the eventual posterior model that corresponds to conscious experience. In the case of Jeff: not only is the perceptual information related to Jeff's "height" highly precise, the brain will learn to consider "height" as a highly precise prior for recognizing Jeff. In a more general way, precision control is of major importance in order to create adaptive models of the stimulation that reaches the brain (Parr & Friston, 2018). Figure 1 illustrates how relative precision of priors and prediction errors (likelihood) impacts the eventual posterior model.

---

Insert Figure 1 here

---

Recently, major advances have been made in relating these predictive processing concepts to a neuroanatomical architecture and in operationalizing them in formal computational ways (see Friston, Rosch, Parr, Price, & Bowman, 2018; Roberts, Friston & Breakspear, 2017; Parr & Friston, 2018).

*2.3. Some implications*

Several implications follow from this way of understanding the functionality of the brain. First, conscious experience always reflects prior expectations to some extent, but the degree to which this happens can vary. Prior beliefs with high precision in the context of imprecise inputs is likely to have a strong impact on the posterior beliefs, whereas the reverse is true when high precision inputs are processed while prior beliefs are imprecise. In both cases, however, the same compelling sense of being "real" or "true" is produced. Second, the criterion for the brain to settle on particular posterior beliefs is not accuracy, but adaptiveness or usefulness. Although accurate beliefs are mostly adaptive, it can be more adaptive to be biased (Lynn & Barrett, 2014). This suggests that it might be more fruitful not to consider bias as mistake or "being wrong" that should

be corrected. Instead, trying to understand its usefulness and modifying conditions to make it less adaptive might be a more fruitful approach to modify the information processing "heuristics". Third, and most importantly, predictive processing does not assume a self or an agent that organizes the information processing traffic. Instead, the brain is considered to self-organize following the principle of free energy minimization and the sense of self or agency is a product of processing heuristics themselves "producing perceiver and percept at the same time" (Friston 2010; Hume, 1739/2007). It follows that N/NE should not be considered something that a self "has", or as some force that impacts and deviates "normal" information processing and should be controlled or regulated by a superseding self. Rather, we suggest to consider N/NE as being inherent to the data processing heuristics themselves (Petersen, von Leupoldt, & Van den Bergh, 2015). The implication is that different characteristics of N/NE should not be seen as separate phenomena that are associated with N/NE, but as representing in some way these very data processing heuristics. In this way, we suggest a more parsimonious interpretation of N/NE and its associated phenomena with potentially important clinical implications. We will expand on this perspective below.

**3. N/NE in a predictive processing model**

Considering N/NE as an adaptive trait involving a recalibration of the threshold for threat detection representing a "better safe than sorry"- strategy, the question is how a BSTS-strategy is conceptualized in a predictive processing model. Because input to the brain as represented in neural distributions can in principle be categorized in an infinite number of patterns, the system has to apply a decision rule to settle on a categorical posterior model (to solve the 'infinity problem', Chater & Vitányi, 2003). Ideally, the decision rule should be set in such a way that important information is not missed, but also that details are not too manifold and/or too specific to impede generalizing and learning from prior experience (Shepard, 1987; Petersen et al., 2015; Rigoli, Pezzulo, Dolan, & Friston, 2017). Threat processing by high N/NE persons can be seen as involving a decision rule that

has shifted towards oversimplifying input. The benefit may be higher speed to categorize input as threat at the expense of the level of detail by which the input is processed.

Applying a decision rule that reduces detailed processing of the prediction errors while maintaining highly precise threat-related priors results in perceptions that are more informed by categorical threat-related priors than by actual input. This process (a form of "jumping to conclusions") has the advantage of providing reduction of uncertainty on the short term, but the cost is that prediction errors tend to remain imprecise, reducing the evidence to update the prior beliefs and enabling high-level (threat-related) priors to further dominate one's conscious experiences. Reduced detail in processing information from the inner and outer world and poor updating of prior beliefs will also result in generative models with low level of detail which, according to computational simulations, increases the precision of predictions (Kwisthout, Bekkering, & Van Rooij, 2017), leading to chronic conditions of uncertainty/surprise on the longer term (i.e. always new unpredicted input will have to be dealt with). This suggest that a "stagnated error reduction process" is at the core of high N/NE[1], involving persistent deviations between model-based prior expectations and actual evidence which is "the hallmark of a bad model" (Stephan et al., 2016; p. 6).

It is important to note that in a predictive coding perspective, N/NE is not just an affective quality added to the error processing dynamics. Rather, emotional valence is considered to emerge from these dynamics at work (Van de Cruys, 2017[2]; Joffily & Coricelli, 2013). For example, unresolved mismatch ('surprise'[3]) between predicted and actual stimulation may characterize feelings of curiosity and interest and generate positive feelings of mastery as long as predictive progress is being made (Van de Cruys, 2017). However, when predictive progress stagnates, the persistent deviations between model-based prior expectations and evidence may engender unproductive coping

1 In line with GUTS (see above) that assumes that expecting threat is actually the default state unless safety is perceived that inhibits it, high N/NE persons can be conceived of as being guided by stronger prior threat-related beliefs, while reduced processing of actual stimuli and context impedes perceiving safety.
2 See Van de Cruys (2017) for an elaborate essay on the conceptualization of affect in a predictive processing view.
3 Violations of expectations ("surprise") should not be seen as "agent level surprise" (the way individuals would consciously experience it), but as sub-personal neural processing products.

behaviors (e.g. chronically elevated vigilance) and inappropriate physiological activations that characterize N/NE, which in turn may further compromise updating of non-adaptive prior beliefs (Stephan et al., 2016).

How this processing strategy is implemented in the brain and how it maps onto amygdala overactivation and deficient cortical control that characterizes N/NE as described above, is currently unclear. In any case, a predictive processing account goes beyond a focus on particular brain structures and considers the process of descending predictions, ascending prediction errors and error minimization across hierarchical layers of the brain as involving information flow through large-scale functional networks (see for examples Barrett & Simmons, 2015; Barrett, 2017; Park & Friston, 2013). How functionality emerges from the structural architecture of the brain still remains to a large extent "a mystery in neuroscience" (Park & Friston, 2013, p. 1). Studies on how N/NE is implemented in large-scale functional neural networks are emerging but, the picture up to now is fragmented by taking a focus on specific aspects of N/NE (e.g. working memory in N/NE persons), different brain connectivity analysis methods and metrics and results are little consistent (see Gentili et al., 2017; Dima et al., 2015; Ueda et al., 2018; Li et al., 2018).

In sum, we suggest that high trait N/NE as a general vulnerability factor for psychopathology has a BSTS-strategy at its core that implies a stagnated error reduction process when processing input. Input is processed with low detail resulting in little precise prediction errors allowing prior threat-related beliefs to dominate the immediate experience. However, on the longer run it also leads to poor updating of prior beliefs, maintaining highly precise prior beliefs and thus to chronic uncertainty/surprise. We suggest that this pervasive way of processing input underlies a wide variety of disparate phenomena and that the phenomenal differences result from the different content of information rather than involving different processes. This is not to say that the phenomena described below must always show up together within one individual. Context and individual concerns and experiences can moderate which of the phenomena below will be more pronounced in a particular case (Nolen-Hoeksema & Watkins, 2011; Lahey et al., 2017; Barlow et al., 2014b).

**4. Exemplars of a general better-safe-than-sorry processing strategy**

Several cognitive biases related to N/NE have been extensively documented and – occasionally - interpreted as examples of a BSTS-strategy (Williams, Mathews & MacLeod, 1996; Mathews & MacLeod, 2005; Cisler & Koster, 2010; Forbes, Purkis, & Lipp, 2011). Typical examples are attentional bias towards threat detection and towards negative emotional information, and more difficulty to disengage from it. Also, bias towards negative interpretation of meanings and to draw negatively valenced inferences has been well documented (Mathews, Ridgeway, Cook & Yiend, 2007; Heinrichs & Hofmann, 2001). Negatively biased recall of information is also a robust finding, particularly when the information relates to the self (Rusting, 1998). Because it is rather obvious that attentional and interpretational biases are examples of lowering the threshold for threat detection, these cognitive biases will not be reviewed here, although they have not yet been conceptualized within a predictive processing framework. In the following, we chose to address other N/NE-related psychological characteristics that are less obvious examples of a BSTS-strategy and discuss how a predictive processing perspective reveals the BSTS-strategy at its core.

*4.1. Perseverative cognition*

Perseverative cognition, or the repetitive, sustained activation of cognitive representations of past stressful events or feared events in the future, refers to a class of cognitive activities subsuming depressive rumination and anxious worry that is both prospectively and cross-sectionally associated with anxiety disorders and depression (Brosschot, Gerin & Thayer, 2006; Drost, Van der Does, van Hemert, Penninx, & Spinhoven, 2014). It can therefore be considered a transdiagnostic feature of internalizing disorders. Importantly, perseverative cognition does not directly cause depression and anxiety but it acts as a mediator between N/NE and these disorders, enhancing their probability. Perseverative cognition is also associated with increased allostatic load on the cardiovascular (blood pressure, heart rate, heart rate variability) and the endocrine system (e.g. cortisol) (Ottaviani et al., 2016). Perseverative cognition is part of a broader class of repetitive

thinking, that is “thinking attentively, repetitively or frequently about one’s self and one’s world” (Watkins, 2008, p. 163), but the maladaptive variant typically has a negatively valent content. Apparently, perseverative cognition is triggered by negative affective states, more precisely when such states are elicited by the awareness of a difference between the current state and a target state (Smith & Alloy, 2009). For example, depressive rumination can be seen as a cognitive elaboration of one’s current sad state and its potential negative consequences in order to attain a more desired state, as well as an elaboration of the reasons for the discrepancy (why-questions).

Interestingly, perseverative cognition is characterized by an abstract level of construal (Watkins, 2008), which is described as forming general, superordinate and decontextualized mental representations that convey the “essential gist and meaning” of events and actions, whereas concrete low-level construals include contextual, specific, and incidental details of events and actions (Watkins, 2008; p. 187). It is assumed that abstract construal is selected as an emotion-focused coping style because it allows cognitive avoidance of a thorough experience of negative affective states. This results in a short-term benefit (avoiding intense negative affect), but also in a long-term problem because more adaptive processing of negative affect and active problem solving is impeded (Smith & Alloy, 2009). The overall effect is prolonged stress and negative affect which promotes a vicious circle with perseverative cognition as the motor that keeps it going.

The idea that perseverative cognition is a chronic attempt to reduce discrepancies between actual and desired goals is central in Control Theory (Martin & Tesser, 1989, 1996), and in an extensive elaboration of it by Watkins (2008). Because both Control Theory and predictive processing are rooted in theories of inference and control in biological systems, it is no wonder that there are striking similarities with the present predictive processing account of N/NE (supra, part 3): The option to select a short-term benefit by avoiding more elaborate processing of negative content at the expense of a larger problem on the longer term represents a BSTS-strategy that is characterized by a stagnated error reduction process. Rather than processing specific and contextualized events and actions that are rich in specific and concrete evidence (prediction errors) and to accommodate

higher order mental representations accordingly (error minimization), abstract construals that largely reflect prior assumptions about the individual in interaction with the world dominate mental activity, compromising adaptiveness of the mental models of oneself and the world on the longer term and leading to chronic uncertainty. As noted by Watkins (2008, p. 192), this strategy is particularly disadvantageous in conditions of novelty, unfamiliarity, difficulty or stress, that is, conditions that are typically threatening for persons with high levels of N/NE.

Although both ways of interpreting perseverative cognition are quite similar, the benefit of understanding perseverative cognition within a predictive coding perspective is that it reveals identical mechanisms that underlie both perseverative cognition and N/NE. This suggests that perseverative cognition is not just a separate phenomenon that is associated with N/NE. Rather, perseverative cognition is at the cognitive level a very instance of the same processing style that constitutes N/NE. Put differently: it is N/NE at work when dealing with discrepancies between actual and desired goals. In this way, a predictive processing account is more parsimonious.

*4.2. Reduced autobiographical memory specificity*

Reduced autobiographical memory specificity is typically assessed with the Autobiographical Memory Test (Williams & Broadbent, 1986; Raes, Hermans, Williams & Eelen, 2007): persons are given an emotional cue-word and are asked to recall specific autobiographical memories in relation to the cue-word. A specific memory refers to a particular event that lasted less than a day, while an overgeneral memory is more generic and typically includes a class of events (categoric memories) or an extended period of time (extended memories) (Griffith et al., 2009). The difficulty to retrieve specific personal memories of a past event is found in a broad range of psychopathological disorders, most importantly depression, post-traumatic stress disorder, but also in acute stress disorder and somatic symptom disorder (Barry, Chiu, Raes, Ricarte & Lau, 2018; Walentynowicz, Raes, Van Diest, & Van den Bergh, 2017). Importantly, reduced autobiographical memory specificity is also a marker of an unfavorable course of psychopathology, affecting severity of symptoms, illness duration, and treatment success (Sumner, Griffith, & Mineka, 2010) and it predicts a gradual increase in depressive

symptoms over the course of 18 months in a community sample (Van Daele, Griffith, Van den Bergh, & Hermans, 2014).

Reduced autobiographical memory specificity is associated with impaired executive function (e.g. deficits in inhibitory control, updating and maintaining information in working memory, verbal fluency) and with avoidance of negative affect (Sumner et al., 2014; Barry et al., 2018). Several brain areas involved in processing emotional salience and self-relevance as well as in executive control, emotion regulation and memory have been associated with reduced autobiographical memory specificity, and these effects may in part be due to (chronic) abnormalities in cortisol. Interestingly, the activation patterns in brain areas are not systematically replicated across diagnostic groups with the same behavioral effects, suggesting that reduced autobiographical memory specificity may come about by a number of different neurocognitive mechanisms (Barry et al., 2018). Possibly, large-scale neural network analysis might extend and further clarify how reduced autobiographical memory specificity is implemented in the brain.

One of the important factors that is assumed to account for reduced autobiographical memory specificity is capture and rumination (CaR-FA-X model; Williams et al., 2007). Capture and rumination imply that memory retrieval thought to be “captured” at a general level, thereby blocking the retrieval of specific memories, while ruminating keeps processing of the information at an analytical and abstract level (see above). Obviously, these deficits make adaptive processing of negative information less likely, leaving one stuck at the level of overgeneral, self-related information. The choice of this level of processing is assumed to result from functional avoidance of episodes of negative affect that might be triggered by emotional cues. On the long run, however, this strategy is maladaptive because it inhibits emotional processing and active problem solving related to the sources of the negative affect. An unresolved issue is whether this strategy already plays at the encoding phase, resulting in poor specific and detailed memory of negatively valent or threatening information in the first place (Raes, 2005).

Within a predictive processing view, this way of processing of information in memory is consistent with the BSTS-strategy described above: little detailed processing of memory input from autobiographical events results in low precise prediction errors, allowing strong prior beliefs representing (negatively valent) generalized memory information to determine conscious recall of self-referential events. Consequently, error minimization stagnates at a rather abstract level leaving generalized prior beliefs unchanged and contributing to a vicious circle of negative affect and low autobiographical memory specificity. Again, when viewed in this way it is clear that reduced autobiographical memory specificity is not just a phenomenon that is associated with N/NE. It is actually the core of N/NE at work when processing of self-relevant memory information.

*4.3. Compromized fear learning*

Associative fear learning typically involves repeated pairing of a (relatively) neutral stimulus with an aversive fear-inducing stimulus (unconditioned stimulus, US), after which the previously neutral, but now conditioned stimulus (CS) elicits fear by itself. Extinction learning implies repeated unreinforced exposures to the CS leading reduced responding to it. Associative fear learning is widely used as a laboratory model to understand pathological fear (Craske, Hermans & Vervliet, 2018). Although strong conclusions about the effect of individual difference variables are hampered by "noise" induced by different methodologies, some evidence suggests specificities in associative fear learning that are associated with N/NE and the vulnerability to develop and/or suffer from actual pathological anxiety (Lonsdorf & Merz, 2017). First, anxiety-prone persons tend to show poor extinction learning, that is, a less steep decline of fear responses during extinction and/or a higher level of remaining fear after extinction (Shechner, Hong, Britton, Pine, & Fox, 2014). Put differently, for an equal amount of expectancy violations, the CS remains longer categorized as potentially dangerous. Poor extinction learning in the laboratory has been shown to predict the development of anxiety symptoms in real life situations (Guthrie & Bryant, 2006).

Second, anxiety-prone persons show impaired safety learning in a differential learning paradigm. The latter involves the presentation of two neutral stimuli, one of which is paired with the

aversive stimulus (CS+) while the other is never paired with it (CS-), turning the neutral stimuli into, respectively, a danger and safety cue. Impaired safety learning, emerging as elevated fear responding to the CS-, has frequently been observed in persons scoring high on trait anxiety (Gazendam, Kamphuis, & Kindt, 2013; Kindt and Soeter, 2014), in persons with subclinical levels of anxiety (Chan and Lovibond, 1996; Haddad, Pritchett, Lissek, & Lau, 2012; Arnaudova et al., 2013) and in persons with anxiety disorders (Jovanovic et al., 2013; Winslow, Noble, & Davis, 2008; Grillon & Morgan, 1999; Lissek et al., 2009) and depression (Pollak et al., 2008). Other data showed that patients with panic disorder had higher danger expectancy when safety stimuli were presented (Lenaert, Boddez, Vervliet, Schruers, & Hermans, 2015). Poor danger-safety discrimination also predicts return of fear after treatment (Staples-Bradley, Treanor & Craske, 2018). In other words, persons vulnerable for or with affective psychopathology show reduced discrimination between danger and safety in a danger context, illustrating a bias towards categorizing a safe cue as dangerous (see also Garcia & Zoellner, 2017).

Third, also overgeneralization of fear has been observed in threat sensitive persons. Generalization occurs when fear has been learned in response to one stimulus and subsequently emerges in response to stimuli that are similar but have never been paired with the aversive stimulus. Overgeneralization, then, occurs when fear responding to generalization stimuli remains relatively higher compared to healthy controls across a decreasing gradient of similarity with the original CS. This suggests that threat sensitive persons are less sensitive to differences between a new stimulus and danger cues leading to a higher probability to categorize them as dangerous. Overgeneralization of learned fear has been found in several groups with anxiety disorders (Lissek, Rabin, Heller, Lukenbaugh, Geraci, Pine, & Grillon, 2010; Lissek, Kaczkurkin, Rabin, Geraci, Pine & Grillon 2014; Lissek & Grillon, 2012; Lenaert et al., 2014).

Fourth, a general sense of unpredictability and uncontrollability contributes to the development of neuroticism. This may result from a lack of safety learning and/or repeated experiences with unpredictable and uncontrollable aversive life events and from harsh, intrusive and

overcontrolling parenting styles. It is associated with alterations in neural function and neuroendocrine regulation of the HPA-stress axis (see Triple Vulnerability Theory, Barlow et al., 2014b).

In summary, persons prone to develop emotional disorders are less likely to develop inhibitory fear learning as indicated by poor extinction, poor safety learning and overgeneralization and they experience the world as less predictable and controllable. Put simply, they are more likely to categorize a safe stimulus as potentially dangerous, they need more expectancy violations to shift a danger cue into a category of safe ones, and – more so than less threat sensitive persons - they tend to categorize both a safe cue within a dangerous context as well as a new cue that resembles a dangerous one, as dangerous.

How do these effects come about according to a predictive processing framework? As suggested above, these effects may result from applying a perceptual decision rule that has shifted towards oversimplifying input (impoverished sampling), allowing highly precise threat-related priors to have stronger impact on the eventual categorical danger perception. The obvious benefit is higher likelihood to categorize input as threat at the expense of the level of detail by which the input is processed. This can be considered an instance of ecologically situated perception (Linson & Friston, 2019). Evidence consistent with this interpretation has been found in relation to fear overgeneralization in anxiety patients. Rather than being a result of post-perceptual choice behavior, it reflects errors resulting from altered perception of simple stimulus features and causing less discrimination between danger and safety cues (Laufer, Israeli & Paz, 2016; see also Struyf, Zaman, Vervliet & Van Diest, 2015; Zaman, Ceulemans, Hermans & Beckers, 2019). Other evidence in support of this interpretation may be found in studies linking these learning abnormalities to memory specificity (above). Accurate discrimination learning relies on sufficiently detailed and specific memory: without detailed memory, experiences will not be represented in memory as unique, specific events reducing the ability to differentiate danger from safety. A study showed that anxiety patients who were low on memory specificity also showed poor discrimination learning in a

differential conditioning paradigm (Lenaert, Boddez, Vervliet, Schruers & Hermans, 2015). It may be also relevant in this context that animal studies have shown that stress-related changes in the hippocampal area lead to deficient "pattern separation" during memory encoding making distinct patterns of information (e.g. CS+ vs CS-) less distinguishable in memory (Tronel, Belnoue, Grosjean, Revest, Piazza, Koehl & Abrous, 2012).

The overall picture of these learning abnormalities suggests that persons high on N/NE use an abstract, general way of construal of sensory input, leaving conscious experience and behavior relatively more influenced by categorical threat-related priors than by actual sensory evidence. Again, this is compatible with a BSTS-strategy when processing aversive information: stimuli are more likely classified as aversive or dangerous at the costs of more errors "on the safe side". This interpretation again suggests that N/NE-related learning abnormalities are not just phenomena associated with N/NE. They actually are the N/NE-related processing heuristics in action during fear learning leading to altered phenomenal learning characteristics.

*4.4. Symptom perception in persons with somatization and medically unexplained symptoms (MUS)*

A pervasive positive relationship has been established between trait N/NE and symptom reports unrelated to physiological dysfunction, or so-called "medically unexplained symptoms" (MUS) (Van den Bergh et al., 2017). Such symptoms are highly prevalent throughout the health care system (Haller, Cramer, Lauche, & Dobos, 2015): they appear in non-consulting individuals where high trait N/NE persons tend to have more symptoms in daily life, in primary care where patients with MUS are characterized by elevated levels of anxiety and depression, and in secondary care where functional somatic disorders (or somatic symptom disorder) are associated with elevated psychiatric comorbidity (Wessely, Nimnuan, & Sharpe, 1999; Henningsen, Zipfel, & Herzog, 2007; Witthöft & Hiller, 2010).

A clue to understand MUS may be found in the fact that neurobiological and psychometric evidence (Walentynowicz, Witthöft, Raes, Van Diest & Van den Bergh, 2018) suggests that the experience of a symptom entails a sensory-perceptual component and an affective-motivational

component that are intuitively integrated into one symptom report. Interestingly, high N/NE persons show enhanced affective-motivational responding to aversive somatic sensations as well as reduced detail in sensory-perceptual processing, blurring a clear distinction between an emotional and a somatic state (Van den Bergh & Walentynowicz, 2016). This results in little precise somatic input (prediction errors) giving room to a large impact of symptom-related categorical priors that are chronically activated in these persons. In other words, in persons with MUS categorical perception of the somatic state is largely determined by somatic priors and less or not by distinct sensory input (Van den Bergh et al., 2017).

Less detailed sensory-perceptual processing of somatic episodes is suggested by a number of recent findings. First, non-consulting high MUS reporters and patients with somatoform disorders show diminished correspondence between induced physiological dysfunction and self-reported symptoms, particularly when in a negative affective context (Van den Bergh, Winters, Devriese, Van Diest, Vos, & De Peuter, 2004; Bogaerts, Notebaert, Van Diest, Devriese, De Peuter, & Van den Bergh 2005). Second, patients with somatoform disorder do not exhibit a peak-end memory bias after an induced aversive somatic episode (Walentynowicz, Bogaerts, Stans, Van Diest, Raes, & Van den Bergh, 2018). The peak-end bias implies that retrospective memory of an event is typically determined by the experience at the peak and at the end. Its absence in retrospective symptom reporting in high habitual symptom reporters and in patients with somatic symptom disorder indicates that their memory of a symptom episode is little affected by the sensory-perceptual changes during a symptom episode. Third, when given health-related cue words for autobiographical memory retrieval, patients with somatic symptom disorder exhibit reduced autobiographical memory specificity after controlling for depression and rumination (Walentynowicz, Raes, Van Diest, & Van den Bergh, 2017). This finding suggests that these patients process and encode health-related episodes in memory in a little detailed way. Fourth, more in-depth analysis of how anxious persons with high habitual symptoms process and categorize ambiguous interoceptive stimuli shows that they more often misclassify interoceptive stimuli of low intensities into a high intensity category

(Petersen, von Leupoldt & Van den Bergh, 2015). Fifth, several impairments in learning to fear pain have been observed in patients with somatic symptom disorder that are in line with the examples mentioned above (see 4.3), such as slower acquisition of pain-related fear to specific cues, more learning of fear of pain in response to the context, more non-differential fear generalization and impaired extinction of generalized fear (Meulders, Jans, & Vlaeyen, 2015; Meulders, Meulders, Stouten, De Bie, & Vlaeyen, 2017; Meulders, Boddez, Blanco, Van Den Houte, & Vlaeyen, 2018). These findings consistently suggest less sensitivity to specific sensory-perceptual evidence and more impact of symptom-related priors in these patients with a somatic symptom disorder.

Other evidence is compatible with the idea that input from somatic and affective states is little distinguished in persons with MUS making them vulnerable for a large impact of somatic priors when in a negative affective state. For example, when inducing negative affect through picture viewing followed by a symptom questionnaire that activates symptom-related priors, elevated symptom reports are elicited in these persons, and this effect is mediated by the difficulty to identify feelings (Van Den Houte et al., 2017). Also, harmless cues that have been associated with symptom episodes in a conditioning paradigm are able to elicit elevated symptoms by themselves. This happens particularly when the cues have a negative affective valence and in persons with high N/NE and in somatic symptom disorder patients (Van den Bergh, Stegen, & Van de Woestijne, 1997; Devriese et al., 2000).

In sum, the above pattern of results converges on the interpretation that persons with high trait N/NE persons are more impacted by the aversive affective-motivational component of a somatic episode, while processing sensory-perceptual information of a somatic episode in a less detailed way. This results in less precise prediction errors allowing a large impact of chronically active symptom-related priors producing symptoms poorly or unrelated to physiological dysfunction. We suggest that this is an instance of a BSTS-strategy: despite little sensory evidence, the brain jumps to the conclusion of an aversive somatic symptom being present. The benefit is a reduction of uncertainty at an early stage of the type “better the devil you know than the one you don’t” (Van

den Bergh et al., 2017; p. 197). The downside is, however, little sensitivity to error correction by means of an update of the prior expectation leading to persistent physical symptoms unrelated to actual physiological dysfunction (Henningsen, Gündel, Kop, Löwe, Martin, Rief, et al., 2018). Again, MUS are in this respect not a separate phenomenon that is associated with high N/NE. It actually is the N/NE-related processing heuristics at work when processing somatic information.

**5. Integration**

Using a predictive processing perspective, we interpreted four behavioral phenomena that are substantially associated with N/NE as exemplars of a BSTS-strategy. We chose these phenomena because, unlike "classic" attentional and misinterpretation biases, they are not typically considered straightforward examples of such a strategy. However, looking at these phenomena using a predictive processing perspective suggests that this strategy may also underlie these examples, allowing a more parsimonious interpretation of a larger range of behavioral facets of N/NE. Indeed, each of these examples can be reframed as emanating from the same deeply embedded processing heuristic that constitute N/NE and produce the conscious experience of the person in the world (Petersen et al., 2015). We propose that this processing heuristic of persons with N/NE gives more weight (precision weighing) to the affective-motivational aspects of the input at the expense of detailed sensory-perceptual processing[4]. Depending on the source and focus of threat, this general processing strategy finds its expression in different behavioral phenotypes: perseverative cognition shows up in relation to information in working memory, reduced autobiographical memory specificity in relation to information in long term memory, while in compromized fear learning and in somatization information from, respectively, extero- and interoceptive threat is at stake.

[4] An interesting conflicting view is offered by researchers promoting the affect-as-information model (Schwarz & Clore, 2003; Lerner, Valdesolo, & Kassam, 2015). This view puts forward that negative emotions signaling threat demand attention and promote vigilant and detailed processing, whereas positive emotions signal safety which would lead to less detailed more heuristic processing. This view would predict a more detailed processing style during negative affect. However, most of this research is based on induction of brief states of sadness in unselected individuals. Therefore, we believe this line of research is less relevant for more permanent and pervasive threat-related information processing.

The obvious short-term benefit of this strategy goes along with long-term costs in the form of a "stagnated error reduction process": little detailed sensory-perceptual processing results in low precise prediction errors giving room to precise threat-related priors to dominate conscious experience and to insufficient error minimization by updating the priors. Conversely, active inference (see above) leads perceptual systems to sample threat-related information consistent with prior expectations. This keeps the system in a chronic state of stress produced by expecting and perceiving threat with little chance for corrective experiences[5].

It is obvious that the term "strategy" should not be considered the result of a deliberate decision by a conscious self to avoid negative states, but a more fundamental way to handle information that in some way appears threatening. This psychobiological strategy is genetically engraved in a substantial part of the population (persons with high trait N/NE) and is epigenetically turned on in persons with early and/or chronic adverse experiences (McEwen et al., 2016). It implies enhanced amygdala function and elevated neurohumoral levels (glucocorticoids, noradrenaline) that are associated with impaired PFC function (Phelps, Lempert & Sokol-Hessner, 2014). Eventually, this strategy may gradually develop from goal-directed to habit-based (e.g. epistemic habit; Friston et al., 2016). This pervasive psychobiological strategy should be considered adaptive in view of both the biological history of our species and the history of individuals with early adverse experiences. Repeated unpredictable and/or chronic threat may have changed both structural and functional properties of the brain to process threat-relevant information in such a way that it promotes the short-term benefit of rapid threat categorization and preparation for action without being outweighed by the less dramatic long-term negative "side-effects".

5 In some ways, the above account resembles the bias towards abstract construals, suggested by Watkins (2011) to characterize patients with (mainly) internalization disorders. It is a challenge for future research to analyze the similarities/differences with our account and to determine whether abstract construals are an equally parsimonious account for the diversity of N/NE-related behavioral phenomena.

## 6. Clinical implications

Considering N/NE as a transdiagnostic trait variable that implies both elevated vulnerability for and a psychobiological dimension underlying internalizing psychopathologies, important questions are how malleable the trait is, whether the trait and its consequences can be prevented and treated and whether this would be a more proficient strategy than targeting specific disorders and symptoms. Overall, it seems that N/NE is more malleable than generally believed and slowly changes over time. Both in children and adults significant small to moderate changes in N/NE were observed as a result of CBT-based interventions targeting N/NE, and changes in N/NE predicted changes in symptoms and functional impairment but not vice versa (Barlow et al., 2014a; Sauer-Zavala, Wilner, & Barlow, 2017; Farchione, Fairholme, Ellard et al. , 2012; Carl, Gallagher, Sauer-Zavala, Bentley, & Barlow, 2014). Recent meta-analytic evidence provides more solid evidence suggesting that N/NE can be altered by treatment with moderate effect size (Cohen's d=.50 -.60). Interestingly, the type of therapy did not matter much, and change reached its maximum effects between 4 and 8 weeks (Roberts et al., 2017).

Although these findings are promising, it remains important to know which components are critical and why, and to design focused intervention protocols to further improve treatment effects. The Unified Protocol for Transdiagnostic Treatment of Emotional Disorders was developed by Barlow and coworkers tries to do this (Ellard, Fairholme, Boisseau, Farchone, & Barlow, 2010; Bullis, Fortune, Farchione, & Barlow, 2014; Ehrenreich-May et al., 2017). Conceiving N/NE as characterized by aversive reactions to emotional experiences and attempts to avoid and escape them, the treatment modules focus on extinguishing distress in response to strong emotions and reduction of avoidant emotion regulation strategies. The latter comes close to what would be needed within the current predictive processing perspective. Assuming that N/NE reflects a stagnated error reduction process, the goal then is to guide persons to overcome the point where error reduction gets stuck. This implies processing information that is associated with aversiveness thoroughly and with openness for detailed sensory-perceptual threat-related elements. This will enable prediction errors to modify

threat-related priors and, on the longer term, facilitate a more complete error minimization process. Once modified, functionally more adaptive priors will in turn act as predictions for new input allowing more efficient error minimization and explaining away unresolved mismatch between predicted and actual input. In other words, a more adaptive generative model about the world is formed, characterized by less unresolved mismatch between expected and actual input.

However, high-level threat-related priors that are innate or learned to fit a survival goal may be quite strong and not easily changed. A predictive processing framework considers priors as forward models, that is, "embodied, whole brain representations" that anticipate upcoming sensory events as well as the best action to deal with the impending sensory events (Barrett, 2017; Barsalou, 2008). This means that they involve internal activation of encodings of actual perceptual, somatovisceral and motor activity that promote prediction error minimization by initiating responses that confirm predictions (active inference). These responses can be unpacked into a large array of lower level component predictions that eventually initiate defensive action programs, including proprioceptive, autonomic and somatovisceral responses associated with defensive activation (Van de Cruys, 2017). We suggest that treatment interventions should impact the way of processing of threat-relevant information at all levels of the "machinery", including altering defensive action programs that are intrinsically part of the prior expectations. This emphasizes the necessity to coach persons towards openness to process aversive information by including components that help the person to disengage from defensive action tendencies.

The above description is consistent with exposure-based interventions if conceived in a different way. While exposure in CBT-accounts relies on an extinction rationale implying a "belief change" (e.g. that a feared stimulus is not coming and/or is not as bad as expected; Foa & McLean, 2016), in the present analysis it is not so much the information that a feared stimulus is not coming that makes the change. If defined as an attentive, open and non-defensive way of processing threat-relevant information, it implies changing the fundamental attitude that characterizes the existence of high N/NE-persons in the world. In other words, it is a fundamental disengagement from defensive

response mobilization during processing of the aversive information, thereby changing associated psychophysiological and motor response programs. Putting it this way turns the sequence in some way upside down: defensive response mobilization is not just something that automatically stops after experiencing that a dreadful stimulus is not coming. It is the disengagement from defensive action tendencies during processing of the stimulus that makes is less dreadful: It means that the weight of the threat-related prior has been importantly reduced and that more weight is given to processing input in an open, non-defensive way. More detailed sensory processing will then promote updating the prior beliefs. This means that information processing is not inspired anymore by a BSTS-strategy, but by a "wait and see" rationale. Interestingly, this reasoning implies that the threatening aspect and aversive impact of any kind of stimulus can possibly be reduced by this approach, regardless whether it is a conditioned or unconditioned source of aversiveness[6].

The critical element advanced here - releasing stagnated error reduction by helping persons to disengage from defensive action tendencies during processing of aversive information – involves a fundamental change in psychophysiological response set to process potentially threatening information. This analysis has two important consequences. First, it suggests the prediction that any treatment should include this critical element to be effective and that sharpening and reinforcing operationalizations of this element will further improve treatment effectiveness. Second, it allows to detect a common ground for many treatment strategies that aim to change habitual information processing of negatively valent information. For example, Concreteness Training (Watkins et al., 2012) aims at changing the abstract processing mode in rumination and perseverative cognition and at promoting concrete processing of difficult, upsetting events by focusing on sensory details, on the specific sequence of events and on possible steps to take after the event. MEmory Specificity Training (MEST; Raes, Williams & Hermans, 2009) tries to change reduced autobiographical memory

6 This perspective helps to understand why some terminal persons may peacefully cope with upcoming death, the ultimate, hardwired unconditional aversive stimulus. Paradoxically, this tends to occur when there is no hope anymore and death becomes inevitable. The message of inevitability may help to disengage from further defensive action tendencies, thereby changing fear of dying.

specificity by coaching participants to practice in retrieving specific memories to cue words, delivered across weekly sessions complemented by homework in-between. Evidence suggests that MEST can improve autobiographical memory performance and cause subsequent reduction in depressive symptoms (Hitchcock, Werner-Seidler, Blackwell, & Dalgleish, 2017). In a similar vein, interoceptive differentiation training has been coined as a treatment strategy to guide patients with medically unexplained symptoms to become more sensitive to sensory-perceptual details of somatic sensations as a way to update chronically active symptom-related priors. Initial evidence suggests, for example, that heartbeat perception training reduces somatization processes (Schaefer et al., 2014). In some way, the idea of making room for more detailed sensory processing without defensive response mobilisation is also captured by the concept of "acceptance" and "self-compassion" (Forman & Herbert, 2009; Gilbert, 2009) and is consistent with a general overarching account of therapeutic change in psychotherapy (Lane, Ryan, Nadel, & Greenberg, 2015).

Finally, also psychopharmacological treatment with serotonergic drugs (SSRI's; selective serotonin reuptake inhibitors) can reduce behavioral manifestations of trait N/NE in healthy persons (Ilieva, 2015). This is not surprising given serotonin's intricate involvement in aversive affective processing and control, also revealed by manipulations such as tryptophan depletion studies in humans (e.g. Ruhé et al., 2007). However, its role is complex as shown by both positive and negative covariances between serotonin and aversion, and its influence on neural plasticity and learning rates. Interestingly, this complexity has led researchers to use computational models to understand the regulatory effects of serotonin, suggesting that it is a signal associated with predictions and prediction errors for future aversive outcomes (Dayan & Huys, 2009; Ilgaya et al., 2018). Here serotonin is particularly implicated in the coding of loss-related prediction errors, thereby inhibiting over-reactions to negative outcomes and modulating behavioral choice selection to handle risk following negative events (Moran et al., 2018). Correspondingly, SSRI's are suggested to boost behavioral treatments by impacting the learning rate and increase underlying neural plasticity to deal

better with (potential) loss and aversion (Iigaya et al., 2018) or, in other words, by releasing stagnated error processing.

## 7. Summary and future directions

Evidence shows that the variation in psychopathology can be parsimoniously modeled using a limited number of hierarchically arranged dimensions. Neuroticism/negative emotionality (N/NE) is such a dimensional trait that appears as a large general vulnerability factor for psychopathology, particularly within the internalization spectrum (Conway et al., 2019). An important question is how to conceive of this trait. Any mechanistic interpretation should meet two important criteria: (1) it should be as parsimonious as possible; (2) it should be able to explain a large variety of phenomenally distinct features of psychopathology. For this exercise we turned to a predictive processing perspective, which conceives of the brain as an active prediction testing organ that tries to make sense of the stimulation it receives. Phenomenal reality emerges from two counterflowing streams of information: expectations (predictions or priors) and prediction errors (input). Prediction errors are propagated through the hierarchical processing architecture of the brain in a prediction error minimization process to eventually settle on a generative model that best explains the input. Both priors and prediction errors should be seen as neural distributions with a variance (precision) that determines the relative weight in the eventual generative model. Within this perspective, we described trait N/NE as reflecting a stagnated error reduction process. We suggested that this results from a generalized BSTS-strategy to process information that is associated with aversiveness: strong threat-related priors prompting active inference and low detailed sensory-perceptual and meaning analysis of the input facilitates fast categorization of threat-related stimuli at the expense of updating threat-related priors. While the resulting generative model may be adaptive on the short term, there is no room to develop a more adaptive generative model of the input on the longer term. We argued that this generalized BSTS-strategy underlies different cognitive and affective features and risk factors of psychopathology that are typically investigated in separate research lines. Besides

more “classic” BSTS exemplars such as attentional and interpretational biases, these include maladaptive perseverative cognition, reduced autobiographical memory, compromized fear learning, and poor symptom processing in somatization. Depending on contextual variables and individual concerns, some high N/NE persons will be characterized more or less by these behavioral features. Based on this analysis, we suggested that a critical element in all treatment approaches is some kind of exposure, defined as an attentive, open and non-defensive way of processing threat-relevant information. We suggested that further and more powerful operationalizations of this critical element when processing specific concerns and fears may be a fruitful way to go in order to dampen trait N/NE and its consequence.

Obviously, the above account implies major challenges for research to test and validate its claims. Predictive processing, being a (Bayesian) computational framework, has stimulated interest in (and calls for) computational psychiatry/psychopathology and psychosomatics (Petzschner, 2017; Petzschner, Weber, Gard, & Stephan, 2017). So, one way to validate its claims is to flesh out a computational version of the predictive processing model above involving a clear mechanistic description of the critical variables and their interactions, to run simulations and compare the results with evidence from real life. Recently, we started to do this by developing a computational model that accounts for bodily symptoms that maintain a strong, weak or absent relationship with bodily input (Pezzulo et al., 2019; Pezzulo, Maisto, Barca, & Van den Bergh, submitted; see osf.io/dywfs). Obviously, this should be done also for other phenomena related to N/NE and subsequently tested with real data. Another approach is to experimentally test specific predictions made by the present account. This may require the development of new paradigms that manipulate both categorical prior expectations and actual input to investigate the relative impact of priors upon the eventual perception and subsequent cognitive processing (see examples related to interoception and symptom perception: Petersen, Schroijen, Mölders, Zenker & Van den Bergh, 2014; Van den Houte et al., 2017; Zacharioudakis, Vlemincx & Van den Bergh, 2020). These are only a few examples that

may point a way to go and a huge challenge remains ahead. However, as Barrett (2017) writes: “At the beginning, new paradigms raise more questions than they answer”.

**8. References**

Arnaudova, I., Krypotos, A. M., Effting, M., Boddez, Y., Kindt, M., & Beckers, T. (2013). Individual differences in discriminatory fear learning under conditions of ambiguity: A vulnerability factor for anxiety disorders? Frontiers in Psychology, 4, 298. doi: 10.3389/fpsyg.2013.00298

Baker, L. A., Cesa, I. L., Gatz, M., & Mellins, C. (1992). Genetic and environmental influences on positive and negative affect: Support for a two-factor theory. Psychology and Aging, 7(1), 158.

Barlow, D. H., Allen, L. B., & Choate, M. L. (2004). Toward a unified treatment for emotional disorders. Behavior Therapy, 35, 205-230. http://dx.doi.org/10.1016/S0005-7894(04)80036-4

Barlow, D. H., Ellard, K. K., Sauer-Zavala, S., Bullis, J. R., & Carl, J. R. (2014, b). The origins of neuroticism. Perspectives on Psychological Science, 9, 481-496. http://dx.doi.org/10.1177/1745691614544528

Barlow, D. H., Sauer-Zavala, S., Carl, J. R., Bullis, J. R., & Ellard, K. K. (2014, a). The nature, diagnosis, and treatment of neuroticism: Back to the future. Clinical Psychological Science, 2, 344-365. http://dx.doi.org/10.1177/2167702613505532

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. Social cognitive and affective neuroscience, 12(1), 1-23. doi.org/10.1093/scan/nsw154

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. Nature Reviews Neuroscience. 16, 1-11. doi:10.1038/nrn3950

Barry, T. J., Chiu, C. P., Raes, F., Ricarte, J., & Lau, H. (2018). The neurobiology of reduced autobiographical memory specificity. Trends in Cognitive Sciences, 22(11), 1038-1049. doi.org/10.1016/j.tics.2018.09.001

Barsalou, L. W. (2008). Grounded cognition. Annual Review of Psychololgy, 59, 617-645. doi.org/10.1146/annurev.psych.59.103006.093639

Bogaerts, K., Notebaert, K., Van Diest, I., Devriese, S., De Peuter, S. & Van den Bergh, O. (2005). Accuracy of respiratory symptom perception in different affective contexts. Journal of Psychosomatic Research, 58, 537-543. doi.org/10.1016/j.jpsychores.2004.12.005

Brosschot, J. F., Verkuil, B., & Thayer, J. F. (2016). The default response to uncertainty and the importance of perceived safety in anxiety and stress: An evolution-theoretical perspective. Journal of Anxiety Disorders, 41, 22-34. doi: 10.1016/j.janxdis.2016.04.012

Brosschot, J. F., Verkuil, B., & Thayer, J. F. (2017). Exposed to events that never happen: Generalized unsafety, the default stress response, and prolonged autonomic activity. Neuroscience & Biobehavioral Reviews, 74, 287-296. doi: 10.1016/j.neubiorev.2016.07.019

Brosschot, J. F., Verkuil, B., & Thayer, J. F. (2018). Generalized unsafety theory of stress: Unsafe environments and conditions, and the default stress response. International Journal of Environmental Research and Public Health, 15(3), 464. doi: 10.3390/ijerph15030464

Brosschot, J. F., Gerin, W., & Thayer, J. F. (2006). The perseverative cognition hypothesis: A review of worry, prolonged stress-related physiological activation, and health. Journal of Psychosomatic Research, 60(2), 113-124.

Brown, T. A., & Barlow, D. H. (2009). A proposal for a dimensional classification system based on the shared features of the DSM-IV anxiety and mood disorders: Implications for assessment and treatment. Psychological Assessment, 21, 256. doi: 10.1037/a0016608

Bullis, J. R., Fortune, M. R., Farchione, T. J., & Barlow, D. H. (2014). A preliminary investigation of the long-term outcome of the Unified Protocol for Transdiagnostic Treatment of Emotional Disorders. Comprehensive Psychiatry, 55, 1920-1927. doi: 10.1016/j.comppsych.2014.07.016

Carl, J. R., Gallagher, M. W., Sauer-Zavala, S. E., Bentley, K. H., & Barlow, D. H. (2014). A preliminary investigation of the effects of the unified protocol on temperament. Comprehensive Psychiatry, 55(6), 1426-1434. doi: 10.1016/j.comppsych.2014.04.015

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., ... & Moffitt, T. E. (2014). The p factor: one general psychopathology factor in the structure of psychiatric disorders? Clinical Psychological Science, 2, 119-137. doi: 10.1177/2167702613497473

Caspi, A., & Moffitt, T. E. (2018). All for one and one for all: Mental disorders in one dimension. American Journal of Psychiatry, 175(9), 831-844. doi: 10.1176/appi.ajp.2018.17121383

Cisler, J. M., & Koster, E. H. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. Clinical Psychology Review, 30, 203-216. doi: 10.1016/j.cpr.2009.11.003

Chan, C. K., & Lovibond, P. F. (1996). Expectancy bias in trait anxiety. Journal of Abnormal Psychology, 105, 637.

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? Trends in Cognitive Sciences, 7, 19-22.

Clark, L. A. (2009). Stability and change in personality disorder. Current Directions in Psychological Science, 18, 27-31. https://doi.org/10.1111/j.1467-8721.2009.01600.x

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36, 181-204. doi: 10.1017/S0140525X12000477

Conway, C. C., Forbes, M. K., Forbush, K. T., Fried, E. I., Hallquist, M. N., Kotov, R., ... & Sunderland, M. (2019). A hierarchical taxonomy of psychopathology can transform mental health research. Perspectives on psychological science, 14(3), 419-436. doi: 10.1177/174569161881069

Cousijn, H., Rijpkema, M., Qin, S., van Marle, H. J., Franke, B., Hermans, E. J., ... & Fernández, G. (2010). Acute stress modulates genotype effects on amygdala processing in humans. Proceedings of the National Academy of Sciences, 107(21), 9867-9872. doi.org/10.1073/pnas.1003514107

Craske, M. G., Hermans, D., & Vervliet, B. (2018). State-of-the-art and future directions for extinction as a translational model for fear and anxiety. Philosophical Transactions of the Royal Society B: Biological Sciences, 373, 20170025. https://doi.org/10.1098/rstb.2017.0025 .

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. BMC Medicine, 11, 126. http:// dx.doi.org/10.1186/1741-7015-11-126

Cuijpers, P., Smit, F., Penninx, B. W., de Graaf, R., ten Have, M., & Beekman, A. T. (2010). Economic costs of neuroticism: A population-based study. Archives of General Psychiatry, 67(10), 1086-1093. doi: 10.1001/archgenpsychiatry.2010.130.

Depue, R. A. (2009). A multidimensional neurobehavioral model of personality disorders. In S. J. Wood, N. B. Allen, & C. Pantelis (Eds.). (2009). The neuropsychology of mental illness. (pp. 300–315). Cambridge, Groot-Brittannië: Cambridge University Press.

Devriese, S., Winters, W., Stegen, K., Van Diest, I., Veulemans, H., Nemery, B., Eelen, P., Van de Woestijne, K., & Van den Bergh, O. (2000). Generalization of acquired somatic symptoms in response to odors : A Pavlovian perspective on Multiple Chemical Sensitivity. Psychosomatic Medicine, 62, 751-759.

Dayan, P., & Huys, Q. J. (2009). Serotonin in affective control. Annual Review of Neuroscience, 32, 95-126. doi.org/10.1146/annurev.neuro.051508.135607

Dima, D., Friston, K. J., Stephan, K. E., & Frangou, S. (2015). Neuroticism and conscientiousness respectively constrain and facilitate short-term plasticity within the working memory neural network. Human brain mapping, 36(10), 4158-4163. doi.org/10.1002/hbm.22906

Drost, J., Van der Does, W., van Hemert, A. M., Penninx, B. W., & Spinhoven, P. (2014). Repetitive negative thinking as a transdiagnostic factor in depression and anxiety: A conceptual replication. Behaviour Research and Therapy, 63, 177-183. doi: 10.1016/j.brat.2014.06.004

Ehrenreich-May, J., Rosenfield, D., Queen, A. H., Kennedy, S. M., Remmes, C. S., & Barlow, D. H. (2017). An initial waitlist-controlled trial of the unified protocol for the treatment of emotional disorders in adolescents. Journal of Anxiety Disorders, 46, 46-55. DOI: 10.1016/j.janxdis.2016.10.006

Eisenberg, N., Zhou, Q., Spinrad, T. L., Valiente, C., Fabes, R. A., & Liew, J. (2005). Relations among positive parenting, children's effortful control, and externalizing problems: A three-wave longitudinal study. Child development, 76(5), 1055-1071.

Ehring, T., & Watkins, E. R. (2008). Repetitive negative thinking as a transdiagnostic process. International Journal of Cognitive Therapy, 1(3), 192-205. doi.org/10.1521/ijct.2008.1.3.192

Ellard, K. K., Fairholme, C. P., Boisseau, C. L., Farchione, T. J., & Barlow, D. H. (2010). Unified protocol for the transdiagnostic treatment of emotional disorders: Protocol development and initial outcome data. Cognitive and Behavioral Practice, 17(1), 88-101. doi: 10.1016/j.beth.2012.01.001

Everaerd, D., Klumpers, F., van Wingen, G., Tendolkar, I., & Fernández, G. (2015). Association between neuroticism and amygdala responsivity emerges under stressful conditions. Neuroimage, 112, 218-224. doi.org/10.1016/j.neuroimage.2015.03.014

Eysenck, H. J. (1947). Dimensions of personality. London, UK: Routledge & Kegan.

Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Thompson-Hollands, J., Carl, J. R., ... & Barlow, D. H. (2012). Unified protocol for transdiagnostic treatment of emotional disorders: a randomized controlled trial. Behavior Therapy, 43(3), 666-678. doi: 10.1016/j.beth.2012.01.001.

Foa, E. B., & McLean, C. P. (2016). The efficacy of exposure therapy for anxiety-related disorders and its underlying mechanisms: the case of OCD and PTSD. Annual Review of Clinical Psychology, 12, 1-28. doi.org/10.1146/annurev-clinpsy-021815-093533

Forbes, S. J., Purkis, H. M., & Lipp, O. V. (2011). Better safe than sorry: Simplistic fear-relevant stimuli capture attention. Cognition & Emotion, 25(5), 794-804. doi: 10.1080/02699931.2010.514710.

Forman, E. M., & Herbert, J. D. (2009). New directions in cognitive behavior therapy: Acceptance-based therapies. In W. O'Donohue & J. E. Fisher, (Eds.). General principles and empirically supported techniques of cognitive behavior therapy. (pp 77-101). American Psychological Association.

Friston, K. (2010). The free-energy principle: a unified brain theory? Nature reviews neuroscience, 11(2), 127-138. doi:10.1038/nrn2787

Friston, K. (2012). Prediction, perception and agency. International Journal of Psychophysiology, 83(2), 248-252. doi: 10.1016/j.ijpsycho.2011.11.014

Friston, K. (2013). Life as we know it. Journal of the Royal Society Interface, 10(86), 20130475. http://dx.doi.org/10.1098/rsif.2013.0475

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2016). Active inference and learning. Neuroscience & Biobehavioral Reviews, 68, 862-879. doi.org/10.1016/j.neubiorev.2016.06.022

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: a process theory. Neural computation, 29(1), 1-49. doi.org/10.1162/NECO_a_00912

Friston, K. J., Rosch, R., Parr, T., Price, C., & Bowman, H. (2018). Deep temporal models and active inference. Neuroscience & Biobehavioral Reviews, 90, 486-501. doi.org/10.1016/j.neubiorev.2018.04.004

Garcia, N. M., & Zoellner, L. A. (2017). Fear generalisation in individuals with high neuroticism: Increasing predictability is not necessarily better. Cognition and Emotion, 31(8), 1647-1662. https://doi.org/10.1080/02699931.2016.1259160

Gazendam, F. J., Kamphuis, J. H., & Kindt, M. (2013). Deficient safety learning characterizes high trait anxious individuals. Biological Psychology, 92(2), 342-352. doi: 10.1016/j.biopsycho.2012.11.006

Gentili, C., Cristea, I. A., Ricciardi, E., Vanello, N., Popita, C., David, D., & Pietrini, P. (2017). Not in one metric: Neuroticism modulates different resting state metrics within distinctive brain regions. Behavioural brain research, 327, 34-43. https://doi.org/10.1016/j.bbr.2017.03.031

Gilbert, P. (1998). The evolved basis and adaptive functions of cognitive distortions. Psychology and Psychotherapy: Theory, Research and Practice, 71(4), 447-463. https://doi.org/10.1111/j.2044-8341.1998.tb01002.x

Gilbert, P. (2009). Introducing compassion-focused therapy. Advances in psychiatric treatment, 15(3), 199-208. doi.org/10.1192/apt.bp.107.005264

Gray, J. A., & McNaughton, N. (1996). The neuropsychology of anxiety: Reprise. In Nebraska symposium on motivation (Vol. 43, pp. 61-134). University of Nebraska Press.

Griffith, J. W., Sumner, J. A., Debeer, E., Raes, F., Hermans, D., Mineka, S., ... & Craske, M. G. (2009). An item response theory/confirmatory factor analysis of the Autobiographical Memory Test. Memory, 17(6), 609-623. doi: 10.1080/09658210902939348.

Grillon, C., & Morgan III, C. A. (1999). Fear-potentiated startle conditioning to explicit and contextual cues in Gulf War veterans with posttraumatic stress disorder. Journal of Abnormal Psychology, 108(1), 134.

Guthrie, R. M., & Bryant, R. A. (2006). Extinction learning before trauma and subsequent posttraumatic stress. Psychosomatic medicine, 68(2), 307-311.

Haaker, J., Lonsdorf, T. B., Schümann, D., Menz, M., Brassen, S., Bunzeck, N., ... & Kalisch, R. (2015). Deficient inhibitory processing in trait anxiety: Evidence from context-dependent fear learning, extinction recall and renewal. Biological Psychology, 111, 65-72. doi.org/10.1016/j.biopsycho.2015.07.010

Haddad, A. D., Pritchett, D., Lissek, S., & Lau, J. Y. (2012). Trait anxiety and fear responses to safety cues: Stimulus generalization or sensitization? Journal of Psychopathology and Behavioral Assessment, 34(3), 323-331. DOI 10.1007/s10862-012-9284-7

Haller, H., Cramer, H., Lauche, R., Dobos, G. (2015). Somatoform disorders and medically unexplained symptoms in primary care - a systematic review and meta-analysis of prevalence. Deutsches Aerzteblatt Online. https://doi.org/10.3238/arztebl.2015.0279

Hanson, J. L., Nacewicz, B. M., Sutterer, M. J., Cayo, A. A., Schaefer, S. M., Rudolph, K. D., ... & Davidson, R. J. (2015). Behavioral problems after early life stress: contributions of the hippocampus and amygdala. Biological Psychiatry, 77(4), 314-323. doi: 10.1016/j.biopsych.2014.04.020

Hariri, A. R. (2009). The neurobiology of individual differences in complex behavioral traits. Annual Review of Neuroscience, 32, 225-247. doi: 10.1146/annurev.neuro.051508.135335

Hariri, A. R., & Holmes, A. (2006). Genetics of emotional regulation: the role of the serotonin transporter in neural function. Trends in Cognitive Sciences, 10(4), 182-191.

Hanson, J. L., Hariri, A. R., & Williamson, D. E. (2015). Blunted ventral striatum development in adolescence reflects emotional neglect and predicts depressive symptoms. Biological psychiatry, 78(9), 598-605. doi: 10.1016/j.biopsych.2015.05.010

Heinrichs, N., & Hofmann, S. G. (2001). Information processing in social phobia: A critical review. Clinical Psychology Review, 21(5), 751-770.

Henningsen, P., Zipfel, S., & Herzog, W. (2007). Management of functional somatic syndromes. The Lancet, 369, 946-955. doi:10.1016/S0140-6736(07)60159-7

Henningsen, P., Gündel, H., Kop, W. J., Löwe, B., Martin, A., Rief, W., ... & Van den Bergh, O. (2018). Persistent physical symptoms as perceptual dysregulation: a neuropsychobehavioral model and its clinical implications. Psychosomatic Medicine, 80(5), 422-431. 10.1097/PSY.0000000000000588

Hermans, E. J., Kanen, J. W., Tambini, A., Fernández, G., Davachi, L., & Phelps, E. A. (2017). Persistence of amygdala–hippocampal connectivity and multi-voxel correlation structures during

awake rest after fear learning predicts long-term expression of fear. Cerebral Cortex, 27(5), 3028-3041. doi.org/10.1093/cercor/bhw145

Hitchcock, C., Werner-Seidler, A., Blackwell, S. E., & Dalgleish, T. (2017). Autobiographical episodic memory-based training for the treatment of mood, anxiety and stress-related disorders: A systematic review and meta-analysis. Clinical Psychology Review, 52, 92-107. doi.org/10.1016/j.cpr.2016.12.003

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. Frontiers in Psychology, 3, 96. doi: 10.3389/fpsyg.2012.00096

Hohwy, J. (2013). The predictive mind. Oxford University Press.

Hume, D.A.(1739/2007). A Treatise of human Nature. Reprint edition, D.F.Norton and M.J.Norton (Eds). Oxford:Oxford University Press

Hur, J., Stockbridge, M. D., Fox, A. S., & Shackman, A. J. (2019). Dispositional negativity, cognition, and anxiety disorders: An integrative translational neuroscience framework. Progress in brain research, 247, 375-436. doi.org/10.1016/bs.pbr.2019.03.012

Iigaya, K., Fonseca, M. S., Murakami, M., Mainen, Z. F., & Dayan, P. (2018). An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. Nature Communications, 9(1), 1-10. doi.org/10.1038/s41467-018-04840-2

Ilieva, I. (2015). Enhancement of healthy personality through psychiatric medication: The influence of SSRIs on neuroticism and extraversion. Neuroethics, 8, 127–137. http://dx.doi.org/10.1007/s12152-014-9226-z

Joffily, M., & Coricelli, G. (2013). Emotional valence and the free-energy principle. PLoS Computational Biology, 9(6. https://doi.org/10.1371/journal.pcbi.1003094.

Jovanovic, T., Sakoman, A. J., Kozarić-Kovačić, D., Meštrović, A. H., Duncan, E. J., Davis, M., & Norrholm, S. D. (2013). Acute stress disorder versus chronic posttraumatic stress disorder: inhibition of fear as a function of time since trauma. Depression and Anxiety, 30(3), 217-224. doi: 10.1002/da.21991

Kendler, K. S., Prescott, C. A., Myers, J., & Neale, M. C. (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. Archives of General Psychiatry, 60, 929–937. doi:10.1001/archpsyc.60.9.929\n60/9/929

Kindt, M., & Soeter, M. (2014). Fear inhibition in high trait anxiety. PloS One, 9(1), e86462.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., ... & Eaton, N. R. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): a dimensional alternative to traditional nosologies. Journal of abnormal psychology, 126(4), 454-477. 10.1037/abn0000258

Kuppens, P., Van Mechelen, I., Nezlek, J. B., Dossche, D., & Timmermans, T. (2007). Individual differences in core affect variability and their relationship to personality and psychological adjustment. Emotion, 7(2), 262-274. Doi: 10.1037/1528-3542.7.2.262

Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. Brain and Cognition, 65, 145-168. doi:10.1016/j.bandc.2007.06.007

Kwisthout, J., Bekkering, H., & Van Rooij, I. (2017). To be precise, the details don't matter: on predictive processing, precision, and level of detail of predictions. Brain and cognition, 112, 84-91. doi.org/10.1016/j.bandc.2016.02.008

Lahey, B. B. (2009). Public health significance of neuroticism. American Psychologist, 64, 241–256. doi:10.1037/a0015309

Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is there a general factor of prevalent psychopathology during adulthood? Journal of Abnormal Psychology, 121, 971–977. http://dx.doi.org/10.1037/a0028355

Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D., & Zald, D. H. (2017). A hierarchical causal taxonomy of psychopathology across the life span. Psychological Bulletin, 143(2), 142. doi: 10.1037/bul0000069.

Lane, R. D., Ryan, L., Nadel, L., & Greenberg, L. (2015). Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: New insights from brain science. Behavioral and Brain Sciences, 38, 1-64. doi:10.1017/S0140525X14000041

Lang, P. J., McTeague, L. M., & Bradley, M. M. (2016). RDoC, DSM, and the reflex physiology of fear: A biodimensional analysis of the anxiety disorders spectrum. Psychophysiology, 53(3), 336-347. doi.org/10.1111/psyp.12462

Laufer, O., Israeli, D., & Paz, R. (2016). Behavioral and neural mechanisms of overgeneralization in anxiety. Current Biology, 26(6), 713-722. doi.org/10.1016/j.cub.2016.01.023

Lenaert, B., Boddez, Y., Griffith, J. W., Vervliet, B., Schruers, K., & Hermans, D. (2014). Aversive learning and generalization predict subclinical levels of anxiety: A six-month longitudinal study. Journal of Anxiety Disorders, 28(8), 747-753. doi: 10.1016/j.janxdis.2014.09.006

Lenaert, B., Boddez, Y., Vervliet, B., Schruers, K., & Hermans, D. (2015). Reduced autobiographical memory specificity is associated with impaired discrimination learning in anxiety disorder patients. Frontiers in psychology, 6, 889. doi.org/10.3389/fpsyg.2015.00889

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015) Emotion and decision making. Annual Review of Psychology, 66. htps://doi.org/10.1146/annurev-psych-010213-115043

Li, Q., Yang, G., Li, Z., Qi, Y., Cole, M.W., & Liu, X. (2017). Conflict detection and resolution rely on a combination of common and distinct cognitive control networks. Neuroscience and Biobehavioral Reviews, 83, 123–131. dx.doi.org/10.1016/j.neubiorev.2017.09.032

Li, B. J., Friston, K., Mody, M., Wang, H. N., Lu, H. B., & Hu, D. W. (2018). A brain network model for depression: From symptom understanding to disease intervention. CNS neuroscience & therapeutics, 24(11), 1004-1019. doi.org/10.1111/cns.12998

Linson, A., & Friston, K. (2019). Reframing PTSD for computational psychiatry with the active inference framework. Cognitive neuropsychiatry, 24(5), 347-368. doi.org/10.1080/13546805.2019.1665994

Lissek, S., Rabin, S. J., McDowell, D. J., Dvir, S., Bradford, D. E., Geraci, M., ... & Grillon, C. (2009). Impaired discriminative fear-conditioning resulting from elevated fear responding to learned safety cues among individuals with panic disorder. Behaviour Research and Therapy, 47(2), 111-118. doi: 10.1016/j.brat.2008.10.017

Lissek, S., Rabin, S., Heller, R.E., Lukenbaugh, D., Geraci, M., Pine, D.S., & Grillon, C. (2010). Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. American Journal of Psychiatry, 167, 47–55. doi: 10.1176/appi.ajp.2009.09030410

Lissek, S., Kaczkurkin, A. N., Rabin, S., Geraci, M., Pine, D. S., & Grillon, C. (2014). Generalized anxiety disorder is assoiated with overgeneralization of classically conditioned fear. Biological Psychiatry, 75(11), 909-915. doi: 10.1016/j.biopsych.2013.07.025

Lissek, S., & Grillon, C. (2012). Learning models of PTSD. In The Oxford handbook of traumatic stress disorders (pp. 175-190). New York, USA: Oxford University Press.

Lonsdorf, T. B., & Merz, C. J. (2017). More than just noise: Inter-individual differences in fear acquisition, extinction and return of fear in humans-Biological, experiential, temperamental factors, and methodological pitfalls. Neuroscience & Biobehavioral Reviews, 80, 703-728. doi.org/10.1016/j.neubiorev.2017.07.007

Lupyan, G. & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. Current Directions in Psychological Science, 24, 279-84. doi: 10.1177/0963721415570732

Lynn, S. K., & Barrett, L. F. (2014). “Utilizing” signal detection theory. Psychological Science, 25, 1663-1673. doi: 10.1177/0956797614541991

Mansell, W., Carey, T. A., & Tai, S. J. (2015). Classification of Psychopathology and Unifying Theory the Ingredients of a Darwinian Paradigm Shift in Research Methodology. Psychopathology Review, 2(1), 129-153. doi.org/10.5127/pr.036114

Martin, L. L. & Tesser, A. (1996). Some ruminative thoughts. Advances in social cognition, 9, 1-47.

Martin, L.L. & Tesser, A. (1989). Toward a motivational and structural theory of ruminative thought. In J.S. Uleman & J.A. Bargh (Eds.), Unintended thought (pp. 306-326). New York, NY, US: Guilford Press.

Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. Annual Review of Clinical Psychology, 1, 167-195.

Mathews, A., Ridgeway, V., Cook, E., & Yiend, J. (2007). Inducing a benign interpretational bias reduces trait anxiety. Journal of behavior therapy and experimental psychiatry, 38(2), 225-236. doi: 10.1016/j.jbtep.2006.10.011

McCrae, R. R., & Costa, P. T. (2003). Personality in adulthood: A five-factor theory perspective. New York, USA: Guilford Press.

McEwen, B. S., Nasca, C., & Gray, J. D. (2016). Stress effects on neuronal structure: hippocampus, amygdala, and prefrontal cortex. Neuropsychopharmacology, 41(1), 3. doi: 10.1038/npp.2015.171

McEwen, B. S., & Gianaros, P. J. (2011). Stress and allostasis-induced brain plasticity. Annual Review of Medicin, 62, 431-445. doi: 10.1146/annurev-med-052209-100430

McTeague, L. M., & Lang, P. J. (2012). The anxiety spectrum and the reflex physiology of defense: from circumscribed fear to broad distress. Depression and anxiety, 29(4), 264-281. doi.org/10.1002/da.21891

Meulders, A., Jans, A., & Vlaeyen, J. W. (2015). Differences in pain-related fear acquisition and generalization: An experimental study comparing patients with fibromyalgia and healthy controls. Pain, 156(1), 108-122. doi: 10.1016/j.pain.0000000000000016

Meulders, A., Meulders, M., Stouten, I., De Bie, J., & Vlaeyen, J. W. (2017). Extinction of fear generalization: A comparison between fibromyalgia patients and healthy control participants. The Journal of Pain, 18(1), 79-95. doi: 10.1016/j.jpain.2016.10.004

Meulders, A., Boddez, Y., Blanco, F., Van Den Houte, M., & Vlaeyen, J. W. (2018). Reduced selective learning in patients with fibromyalgia vs healthy controls. Pain, 159(7), 1268-1276. DOI: 10.1097/j.pain.0000000000001207

Moran, R. J., Kishida, K. T., Lohrenz, T., Saez, I., Laxton, A. W., Witcher, M. R., ... & Montague, P. R. (2018). The protective action encoding of serotonin transients in the human brain. Neuropsychopharmacology, 43(6), 1425-1435. doi.org/10.1038/npp.2017.304

Nemeroff, C. (2013). Psychoneuroimmunoendocrinology: The biological basis of mind-body physiology and pathophysiology. Depression and Anxiety, 30, 285–287. doi: 10.1002/ da.22110

Nettle, D., & Bateson, M. (2012). The evolutionary origins of mood and its disorders. Current Biology, 22(17), 712-721. doi: 10.1016/j.cub.2012.06.020

Nolen-Hoeksema, S., & Watkins, E. R. (2011). A heuristic for developing transdiagnostic models of psychopathology: Explaining multifinality and divergent trajectories. Perspectives on Psychological Science, 6(6), 589-609. doi: 10.1177/1745691611419672

Ormel, J., Riese, H., & Rosmalen, J. G. (2012). Interpreting neuroticism scores across the adult life course: immutable or experience-dependent set points of negative affect? Clinical Psychology Review, 32(1), 71-79. doi: 10.1016/j.cpr.2011.10.004.

Ormel, J., Bastiaansen, A., Riese, H., Bos, E. H., Servaas, M., Ellenbogen, M., ... & Aleman, A. (2013). The biological and psychological basis of neuroticism: current status and future directions. Neuroscience & Biobehavioral Reviews, 37(1), 59-72. doi: 10.1016/j.neubiorev.2012.09.004

Ottaviani, C., Thayer, J. F., Verkuil, B., Lonigro, A., Medea, B., Couyoumdjian, A., & Brosschot, J. F. (2016). Physiological concomitants of perseverative cognition: A systematic review and meta-analysis. Psychological Bulletin, 142(3), 231. doi: 10.1037/bul0000036

Parr, T., & Friston, K. J. (2018). The anatomy of inference: Generative models and brain structure. Frontiers in computational neuroscience, 12 (doi: 10.3389/fncom.2018.00090)

Paulus, M. P., & Stein, M. B. (2006). An insular view of anxiety. Biological Psychiatry, 60, 383-387. doi: 10.1016/j.biopsych.2006.03.042

Petersen, S., Schroijen, M., Mölders, C., Zenker, S., & Van den Bergh, O. (2014). Categorical Interoception Perceptual Organization of Sensations From Inside. Psychological Science, 25(5), 1059-1066. https://doi.org/10.1177/0956797613519110

Petersen, S., von Leupoldt, A., & Van den Bergh, O. (2015). Interoception and the uneasiness of the mind: Affect as perceptual style. Frontiers in Psychology, 6, 1408. doi: 10.3389/fpsyg.2015.01408

Petzschner, F. H. (2017). Stochastic Dynamic Models for Computational Psychiatry and Computational Neurology. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2(3), 214-215. /doi.org/10.1016/j.bpsc.2017.03.003

Petzschner, F. H., Weber, L. A., Gard, T., & Stephan, K. E. (2017). Computational psychosomatics and computational psychiatry: toward a joint framework for differential diagnosis. Biological Psychiatry, 82(6), 421-430. doi.org/10.1016/j.biopsych.2017.05.012

Pezzulo, G., Maisto, D., Barca, L., & Van den Bergh, O. (2019). Symptom perception from a predictive processing perspective. Clinical Psychology in Europe. 1(4). doi.org/10.32872/cpe.v1i4.35952

Phelps, E. A., Lempert, K. M., & Sokol-Hessner, P. (2014). Emotion and decision making: multiple modulatory neural circuits. Annual Review of Neuroscience, 37, 263-287.

Pollak, D. D., Monje, F. J., Zuckerman, L., Denny, C. A., Drew, M. R., & Kandel, E. R. (2008). An animal model of a behavioral intervention for depression. Neuron, 60(1), 149-161. doi: 10.1146/annurev-neuro-071013-014119.

Raes, F. (2005). Specificity of autobiographical memory. An experimental investigation of the functional aspects and a prospective investigation of the predictive value for depression. Doctoral thesis, KU Leuven- University of Leuven.

Raes, F., Hermans, D., Williams, J. M. G., & Eelen, P. (2007). A sentence completion procedure as an alternative to the Autobiographical Memory Test for assessing overgeneral memory in non-clinical populations. Memory, 15(5), 495-507.

Raes, F., Williams, J. M. G., & Hermans, D. (2009). Reducing cognitive vulnerability to depression: A preliminary investigation of MEmory Specificity Training (MEST) in inpatients with depressive symptomatology. Journal of behavior therapy and experimental psychiatry, 40(1), 24-38. /doi.org/10.1016/j.jbtep.2008.03.001

Rigoli, F., Pezzulo, G., Dolan, R., & Friston, K. (2017). A Goal-Directed Bayesian Framework for Categorization. Frontiers in psychology, 8, 408. doi.org/10.3389/fpsyg.2017.00408

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. Psychological Bulletin, 143(2), 117-141. Doi: 10.1037/bul0000088

Roberts, J. A., Friston, K. J., & Breakspear, M. (2017). Clinical applications of stochastic dynamic models of the brain, part I: A primer. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2(3), 216-224. doi.org/10.1016/j.bpsc.2017.01.010

Ruhé, H. G., Mason, N. S., & Schene, A. H. (2007). Mood is indirectly related to serotonin, norepinephrine and dopamine levels in humans: a meta-analysis of monoamine depletion studies. *Molecular Psychiatry, 12(4*), 331-359. doi.org/10.1038/sj.mp.4001949

Rusting, C. L., 1998. Personality, mood, and cognitive processing of emotional information: three conceptual frameworks. Psychological Bulletin, 124, 165–196.

Sauer-Zavala, S., Wilner, J. G., & Barlow, D. H. (2017). Addressing neuroticism in psychological treatment. Personality Disorders: Theory, Research, and Treatment, 8(3), 191. http://dx.doi.org/10.1037/per0000224

Schaefer, M., Egloff, B., Gerlach, A. L., & Witthöft, M. (2014). Improving heartbeat perception in patients with medically unexplained symptoms reduces symptom distress. Biological Psychology, 101, 69-76. doi.org/10.1016/j.biopsycho.2014.05.012

Schwarz, N., & Clore, G. L. (2003). Mood as information: 20 years later. Psychological Inquiry, 14(3-4), 296-303. https://doi.org/10.1080/1047840X.2003.9682896

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. Trends in Cognitive Sciences, 17, 565-573. doi: 10.1016/j.tics.2013.09.007

Shackman, A. J., Tromp, D. P., Stockbridge, M. D., Kaplan, C. M., Tillman, R. M., & Fox, A. S. (2016, a). Dispositional negativity: An integrative psychological and neurobiological perspective. Psychological bulletin, 142(12), 1 1275-1314. /dx.doi.org/10.1037/bul0000073

Shackman, A. J., Stockbridge, M. D., Tillman, R. M., Kaplan, C. M., Tromp, D. P., Fox, A. S., & Gamer, M. (2016, b). The neurobiology of dispositional negativity and attentional biases to threat: implications for understanding anxiety disorders in adults and youth. Journal of Experimental Psychopathology, 7(3), 311-342. doi.org/10.5127/jep.054015

Shechner, T., Hong, M., Britton, J. C., Pine, D. S., & Fox, N. A. (2014). Fear conditioning and extinction across development: evidence from human studies and animal models. Biological Psychology, 100, 1-12. doi: 10.1016/j.biopsycho.2014.04.001.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. Science, 237(4820), 1317-1323.

Smith, J. M., & Alloy, L. B. (2009). A roadmap to rumination: A review of the definition, assessment, and conceptualization of this multifaceted construct. Clinical Psychology Review, 29(2), 116-128. doi: 10.1016/j.cpr.2008.10.003

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

Spielberger, C. D. (1966). The effects of anxiety on complex learning and academic achievement. In: C. D. Spielberger (Ed.), Anxiety and behaviour (pp.361-398). New York, USA: AcademicPress.

Spielberger,C.D. (1985). Anxiety, cognition and affect: a state-trait perspective. Hillsdale, NJ: Lawrence Erlbaum Associates.

Staples-Bradley, L. K., Treanor, M., & Craske, M. G. (2018). Discrimination between safe and unsafe stimuli mediates the relationship between trait anxiety and return of fear. Cognition and Emotion, 32(1), 167-173. doi: 10.1080/02699931.2016.1265485

Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A., Paliwal, S., Gard, T., ... & Petzschner, F. H. (2016). Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. Frontiers in human neuroscience, 10, 550. doi.org/10.3389/fnhum.2016.00550

Struyf, D., Zaman, J., Vervliet, B., & Van Diest, I. (2015). Perceptual discrimination in fear generalization: Mechanistic and clinical implications. Neuroscience & Biobehavioral Reviews, 59, 201-207. doi.org/10.1016/j.neubiorev.2015.11.004

Sumner, J. A., Griffith, J. W., & Mineka, S. (2010). Overgeneral autobiographical memory as a predictor of the course of depression: A meta-analysis. Behaviour Research and Therapy, 48(7), 614-625. doi: 10.1016/j.brat.2010.03.013

Sumner, J. A., Mineka, S., Adam, E. K., Craske, M. G., Vrshek-Schallhorn, S., Wolitzky-Taylor, K., & Zinbarg, R. E. (2014). Testing the CaR–FA–X model: Investigating the mechanisms underlying reduced autobiographical memory specificity in individuals with and without a history of depression. Journal of Abnormal Psychology, 123(3), 471. doi: 10.1037/a0037271.

Tronel, S., Belnoue, L., Grosjean, N., Revest, J. M., Piazza, P. V., Koehl, M., & Abrous, D. N. (2012). Adult-born neurons are necessary for extended contextual discrimination. Hippocampus, 22, 292–298. doi:10.1002/hipo.20895

Ueda, I., Kakeda, S., Watanabe, K., Sugimoto, K., Igata, N., Moriya, J., ... & Korogi, Y. (2018). Brain structural connectivity and neuroticism in healthy adults. Scientific reports, 8(1), 16491. doi:10.1038/s41598-018-34846-1

Van Daele, T., Griffith, J. W., Van den Bergh, O., & Hermans, D. (2014). Overgeneral autobiographical memory predicts changes in depression in a community sample. Cognition and Emotion, 28, 1303-1312. doi: 10.1080/02699931.2013.879052

Van de Cruys, S. (2017). Affective value in the predictive mind. In T. Metzinger & W. Wiese (Eds.). Philosophy and Predictive Processing: 24. Frankfurt am Main: MIND Group. doi: 10.15502/9783958573253

Van den Bergh, O., Stegen, K. & Van de Woestijne, K.P. (1997). Learning to have psychosomatic complaints : Conditioning of respiratory behavior and complaints in psychosomatic patients. Psychosomatic Medicine, 59, 13-23.

Van den Bergh, O., Winters, W., Devriese, S., Van Diest, I., Vos, G. & De Peuter, S. (2004). Accuracy of Respiratory Symptom Perception in Persons with High and Low Negative Affectivity. Psychology & Health, 19(2), 213-222. doi.org/10.1080/08870440410001675627

Van den Bergh, O., & Walentynowicz. M. (2016). Accuracy and bias in retrospective symptom reporting. Current Opinion in Psychiatry, 29, 302-308. doi: 10.1097/YCO.0000000000000267

Van den Bergh, O., Witthöft, M., Petersen, S., & Brown, R. J. (2017). Symptoms and the body: taking the inferential leap. Neuroscience & Biobehavioral Reviews, 74, 185-203. doi:10.1016/j.neubiorev.2017.01.015.

Van Den Houte, M., Bogaerts, K., Van Diest, I., De Bie, J., Persoons, Ph., Van Oudenhove, L., & Van den Bergh, O. (2017). Inducing somatic symptoms in functional syndrome patients: Effects of manipulating state negative affect. Psychosomatic Medicine, 79(9), 1000-1007. doi: 10.1097/PSY.0000000000000527

Walentynowicz, M., Raes, F., Van Diest, I., & Van den Bergh, O. (2017). The specificity of health-related autobiographical memories in patients with Somatic Symptom Disorder. Psychosomatic Medicine, 79, 43–49. DOI: 10.1097/PSY.0000000000000357

Walentynowicz, M., Bogaerts, K., Stans, L., Van Diest, I., Raes, F. & Van den Bergh, O (2018). Retrospective memory for symptoms in patients with medically unexplained symptoms. Journal of Psychosomatic Research, 105, 37-44. https://doi.org/10.1016/j.jpsychores.2017.12.006

Walentynowicz, M., Witthöft, M., Raes, F., Van Diest, I., & Van den Bergh, O. (2018). Sensory and affective components of symptom perception: A psychometric approach. Journal of Experimental Psychopathology, 9(2), jep-059716. doi.org/10.5127/jep.059716

Walker, W. R., Yancu, C. N., & Skowronski, J. J. (2014). Trait anxiety reduces affective fading for both positive and negative autobiographical memories. Advances in cognitive psychology, 10(3), 81-89. doi: 10.5709/acp-0159-0

Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. Psychological Bulletin, 134(2), 163. doi: 10.1037/0033-2909.134.2.163.

Watkins, E. (2011). Dysregulation in level of goal and action identification across psychological disorders. Clinical Psychology Review, 31(2), 260-278. doi.org/10.1016/j.cpr.2010.05.004

Watkins, E. R., Taylor, R. S., Byng, R., Baeyens, C., Read, R., Pearson, K., & Watson, L. (2012). Guided self-help concreteness training as an intervention for major depression in primary care: A phase II randomized controlled trial. Psychological Medicine, 42(7), 1359-1371. Doi:10.1017/S0033291711002480

Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. Psychological Bulletin, 96, 465-490. http://dx.doi.org/10.1037/0033-2909.96.3.465

Wessely, S., Nimnuan, C., & Sharpe, M. (1999). Functional somatic syndromes: one or many? The Lancet, 354(9182), 936-939.

Wiese, W., & Metzinger, T. (2017). Vanilla PP for philosophers: A primer on predictive processing. In: T. Metzinger, & W. Wiese (editors). Philosophy and predictive processing: 1 (pp. 8-25). Frankfurt am Main: MIND Group. doi:10.15502/9783958573024.

Williams, J. M., & Broadbent, K. (1986). Autobiographical memory in suicide attempters. Journal of Abnormal Psychology, 95(2), 144-149. http://dx.doi.org/10.1037/0021-843X.95.2.144

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. Psychological Bulletin, 120(1), 3-24.

Williams, J. M. G., Barnhofer, T., Crane, C., Herman, D., Raes, F., Watkins, E., & Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. Psychological Bulletin, 133(1), 122-148.

Winslow, J. T., Noble, P. L., & Davis, M. (2008). AX+/BX− discrimination learning in the fear-potentiated startle paradigm in monkeys. Learning & Memory, 15(2), 63-66. doi: 10.1101/lm.843308

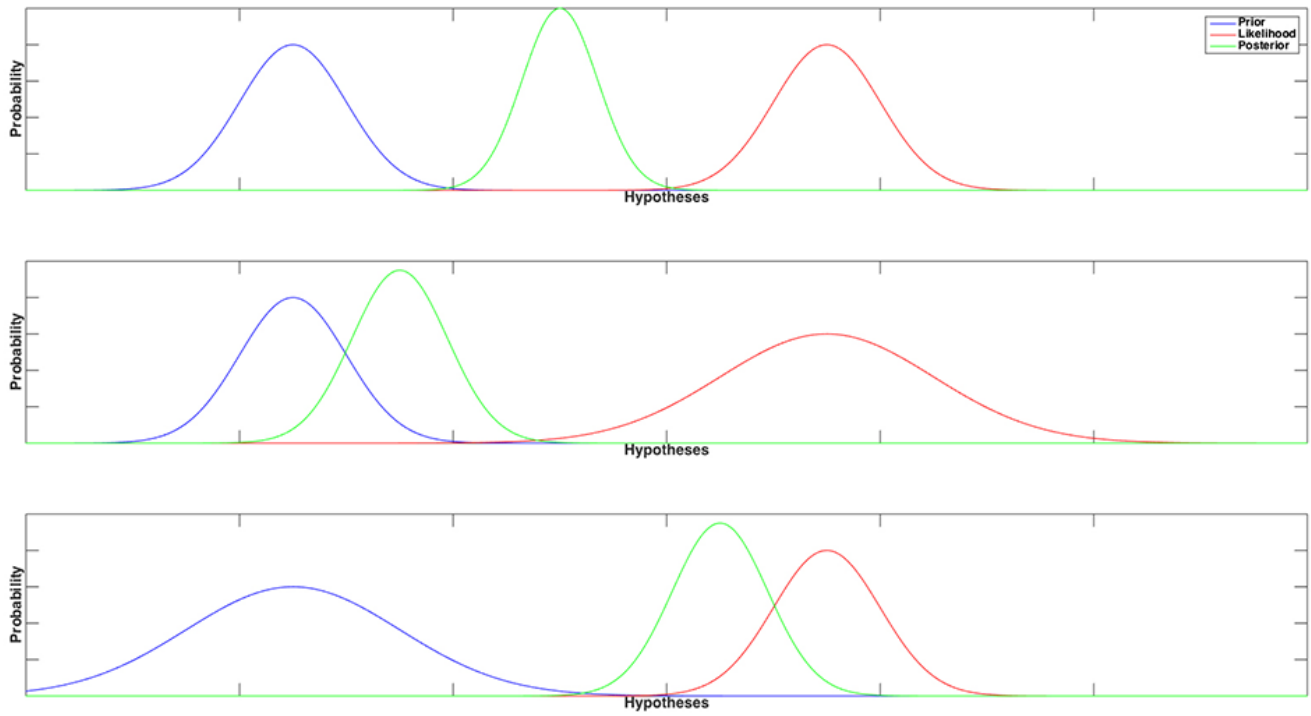
Witthöft, M., & Hiller, W. (2010). Psychological approaches to origins and treatments of somatoform disorders. Annual Review of Clinical Psychology, 6, 257-283. doi: 10.1146/annurev.clinpsy.121208.131505

Zacharioudakis, N., Vlemincx, E., & Van den Bergh, O. (2020). Categorical interoception and the role of threat. International Journal of Psychophysiology. https://doi.org/10.1016/j.ijpsycho.2019.12.009

Zaman, J., Ceulemans, E., Hermans, D., & Beckers, T. (2019). Direct and indirect effects of perception on generalization gradients. Behaviour research and therapy, 114, 44-50. doi.org/10.1016/j.brat.2019.01.006

Zinbarg, R. E., Mineka, S., Bobova, L., Craske, M. G., Vrshek-Schallhorn, S., Griffith, J. W., ... & Anand, D. (2016). Testing a hierarchical model of neuroticism and its cognitive facets: latent structure and prospective prediction of first onsets of anxiety and unipolar mood disorders during 3 years in late adolescence. Clinical Psychological Science, 4(5), 805-824. https://doi.org/10.1177/2167702615618162

Figure 1




*Note*. The top panel shows that if prior and likelihood have the same precision (i.e., inverse variance of the Gaussian distribution), the posterior belief is in between. The second and third panels show that higher precision prior and likelihood "attract" the posterior, respectively. Note that in all cases, the precision of the posterior increases compared to the prior. (From Pezzulo, Maisto, Barca,& Van den Bergh, 2019)