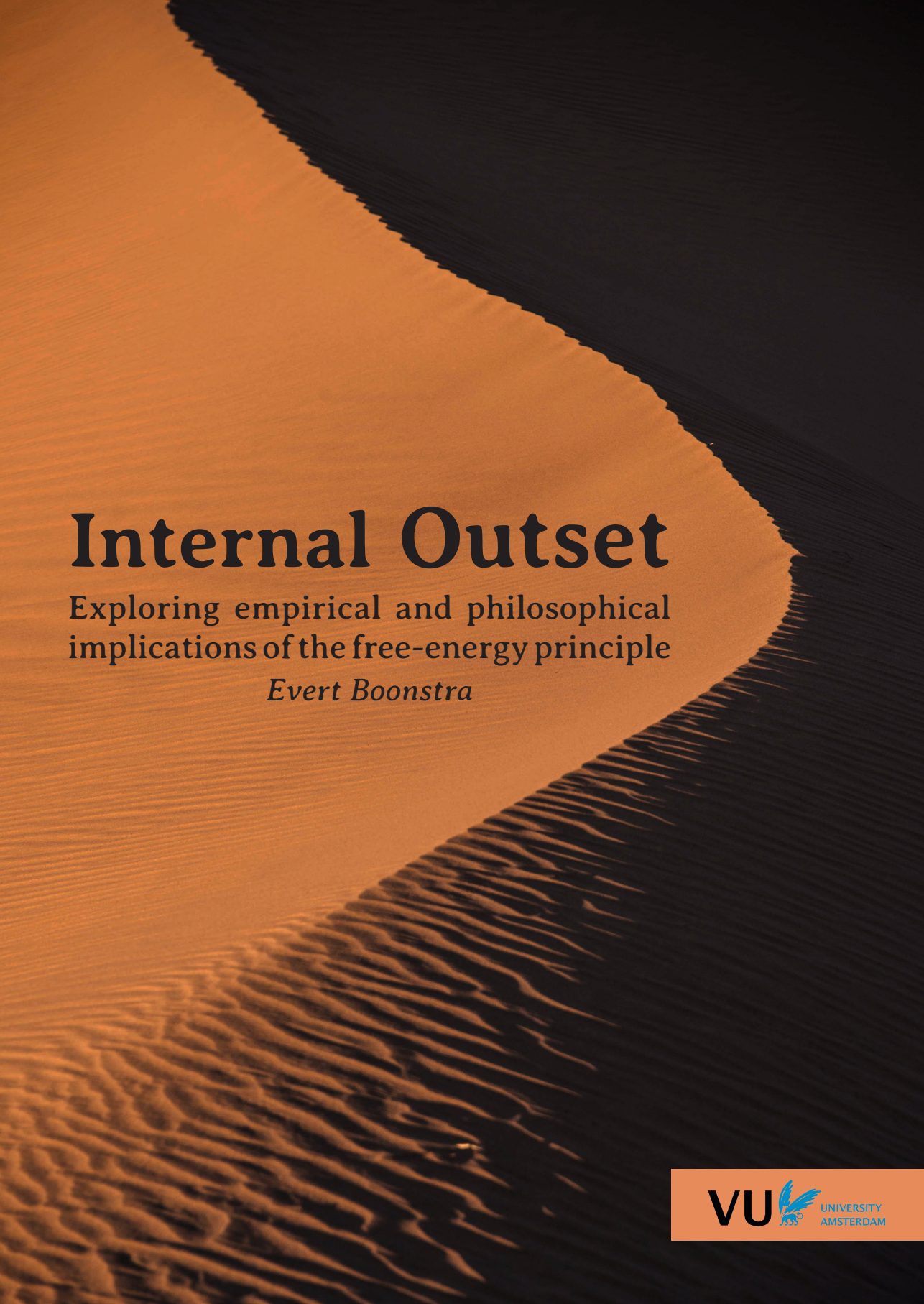
Internal outset

Boonstra, Evert Alexander

2023

DOI (link to publisher)
10.5463/thesis.64

document version
Publisher's PDF, also known as Version of record



Internal Outset

Exploring empirical and philosophical implications of the free-energy principle

Evert Boonstra

VU UNIVERSITY AMSTERDAM

Internal outset

Exploring empirical and philosophical implications of the free-energy principle

Colophon
Printing: Proefschriften.nl, Ede
Layout and design: Birgit Vredenburg, persoonlijkproefschrift.nl

Cover picture
Claudia Gschwend
https://unsplash.com/photos/7afqJT5evIs

Funding
All work presented in this dissertation was supported by the European Research Council (ERC) Starting Grant (ERC-2015-STG-679399) awarded to prof. dr. Heleen A. Slagter

VRIJE UNIVERSITEIT

INTERNAL OUTSET: EXPLORING EMPIRICAL AND PHILOSOPHICAL IMPLICATIONS OF THE FREE-ENERGY PRINCIPLE

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Gedrags- en Bewegingswetenschappen
op woensdag 22 maart 2023 om 13.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

Evert Alexander Boonstra

geboren te Alkmaar

promotoren:	prof.dr. H.A. Slagter prof.dr. C.N.L. Olivers
promotiecommissie:	prof.dr. J.B.J. Smeets prof.dr. B.U. Forstmann prof.dr. F.P. de Lange prof.dr. C. Press dr. J.D. Kiverstein dr. F. van Opstal

Table of contents

Chapter 1

General introduction

Over the past decade or so, the notion that *everything* that the brain does can be explained by one and the same principle has gained scientific interest and traction. This notion goes by many names, such as predictive processing and active inference, but our name of choice here will be the free-energy principle (FEP) (Friston, 2005, 2010, 2019).

On the one hand the FEP has been called tautological, a truism, unfalsifiable, and vacuous. Not only by its critics, but in the first place by its inventor and primary advocate Karl Friston. On the other hand, it has been heralded as a paradigm shift for the field of cognitive neuroscience and psychology. What is this principle and why is it important for the study of brain and behavior?

1.1. Minimal free-energy minimization

Much has been written about the FEP's math and technical details (see Andrews, 2021; Friston, 2010, 2019; Gershman, 2019). What follows will be about the FEP's basic idea instead of math and technicalities to get to the core of its novelty, which provides background to each of the chapters comprising this dissertation.

Boundary

The FEP takes up the question posed by the physicist Erwin Schrödinger: 'How can the events *in space and time* which take place within the spatial boundary of a living organism be accounted for by physics and chemistry?' (Schrodinger, 2012: 3, italics in original; Friston, 2013). A spatial boundary is the FEP's departure point for thinking about biological systems generally and the brain in particular. This boundary instantiates a separation between a system's internal organization and its outside world, or simply between internal and external states. When we speak of a system's "state", we mean simply its coordinates in the space of possible states, with different axes for different variables (Friston, 2018). The FEP takes the the existence of a boundary as its starting point, because if this separation between internal and external states could not be made in spatial terms, there would be no system (no living organism) to speak of or to do research on in the first place. The boundary states that make up the separation between internal and external states in turn consist of sensory states and action states. The relationship between these four types of states (internal, action, external, sensory) render internal states conditionally independent from external states (Friston, 2019), to the point that together they instantiate a feedback loop. For example, say I uphold the (internal) belief that I control the mouse cursor on my computer screen, and I act on this belief, then my internal states affect action states to move the mouse, which in turn affect

external states (the mouse and mouse cursor) that translate back to my sensory states in the form of perceived movement, which would bolster my belief in the control I exert, assuming all goes well.

Attracting set

Now the FEP not only describes a delineation of states, but it formalizes how the system's overall set of states changes over time. Specifically, the FEP formalizes how biological systems maintain their physical integrity by revisiting a small number of characteristic, phenotypic states (Badcock et al., 2019). This is important, because out of all possible configurations a system can take up, only a relatively small number of states can support the continuation of its existence (Wiese & Metzinger, 2017). This limited number of states is formalized as an attracting set of states to which an organism tends to return by minimizing free energy (Friston et al., 2020). For example, most organisms have a narrow range of bodily temperatures allowing for continued existence (~32-42 °C for humans).

Free energy here refers to variational free-energy, which is an information theoretic quantity that bounds or limits the entropy of an organism's sensory states (Friston, 2010). In this context, entropy is a measure of information capturing the long-term average of surprise: a statistical measure of the probability of sensory samples sampled by an agent (Badcock et al., 2019). This means that an organism minimizes variational free-energy by ensuring that it limits the range of states it occupies, thereby increasing its chance of survival.

If a system's state at any given point in time refers to its coordinates in state space, then the change of a system's states over time is the groove it carves through this space of possible states, which involves repeatedly revisiting a small number of states, as in the case of daily routines:

> 'You are you because you revisit (the neighborhood of) these attracting states time after time. Your life traces out a path on this delicately structured attracting set or manifold, where your highly convoluted orbits—or strange loops—keep bringing you back to where you once came from' (Friston, 2018: 3)

Bayesian inference

What makes the FEP such an encompassing framework is that the minimization of free energy or surprise is not only a way for a system to constrain its states to viable bounds, but it is equivalent to the negative logarithm of Bayesian model evidence, which means that the minimization of free energy is equivalent to the optimization of Bayesian model evidence (Friston, 2010). The model in question

coincides with an organism's internal states, a generative model of the (hidden) causes of sensory samples the agent is confronted with, which are encoded by nervous system dynamics, such as neural activity and connection strengths (Badcock et al., 2019). It is important to emphasize that we are not dealing with a brain or nervous system which *also* harbors a model of the causes of its sensations; the idea is that nervous systems instantiate generative models: a biological system *is* a model of its world (Friston, 2010). Said differently, the internal states discussed so far can be cast as encoding Bayesian belief about the state of the (external) world (see Pouget et al., 2013 for possible neural implementations). It is through the equivalence of (sensory) surprise and (Bayesian) negative model evidence that the minimization of free-energy can be cast as the minimization of prediction error; the minimization of the difference between prior beliefs and bottom-up sensory input (Clark, 2013; Hohwy, 2013). To minimize this difference, biological systems tend towards the most likely percept given available sensory data, but they also have the option to change this data through action in line with their current model. Both perception and action are obvious candidates for free-energy minimization at the periphery, but there are more. The brain as a generative model attempts to optimize itself both offline by dreaming (Hobson et al., 2014; Hobson & Friston, 2014, 2016), as well as online by simulating possible courses of action that are not necessarily executed (Seth, 2014, 2015). These are all merely different ways of achieving the same thing: the minimization of free energy or the optimization of the system's generative model given its history and surroundings.

In sum

We can summarize the FEP's basic idea as three central tenets:

1. A boundary between internal states and external states
2. An imperative to revisit a limited set of characteristic states
3. The equivalence of this imperative to Bayesian inference

Taken together, the FEP provides us with an encompassing framework from where to think of biological systems and the brain including the imperative to survive (by occupying characteristic states) on the one hand, and probabilistic inference on the other. With regards to the human brain in particular, perhaps the quickest way to get at the novelty of the FEP is to take the example of hallucinations, where prior beliefs predominate over and above new information (Corlett et al., 2019; Friston et al., 2014). While we usually (and justifiably) ask in what way hallucinations are a departure from normality, with the FEP we get to

turn this question around by asking: what does the possibility of hallucinatory states imply for normal brain function? The answer from the FEP's perspective is that every percept is to a degree a matter of hallucinating at the world. A constitutive boundary between internal belief and outside world inscribes the possibility of false inference in general, and of full-blown hallucination in the case of the human brain, into the very starting point of our approach to understanding the brain.

1.2. Philosophical implications

In contrast to the traditional approach to studying brain and behavior, the FEP's basic idea is nothing less than an alternative overarching theory of biological systems. Scientifically, whether the FEP's view will come to supplant the traditional one will depend on its perspective giving rise to research programs better able at explaining brain function through rigorous experimentation and computational modelling. Only time will tell.

There is however more to say about a theoretical revision of how we understand human brain function. Unlike other fields of science, our understanding of the human brain not only relates to how we think about ourselves and the way we relate to the world outside ourselves, but it also relates to how we think about thinking. Insofar as cognitive neuroscience is a field of research reliant on the thought and brains of individual scientists, insofar as brains are organs performing Bayesian inference, and insofar as Bayesian inference operates based on priors, it seems reasonable to ask: what are our priors as cognitive neuroscientists? Based on what priors does our field of research operate?

Not that this question is new. It is essentially the question posed by Immanuel Kant in his famous three critiques published between 1781 and 1790, but tailored to cognitive neuroscience. For Kant, the problem of philosophy is the observing subject's possibility of knowing, insofar as knowing itself presupposes a faculty or structure that makes knowledge possible. If the aim of science is to gain knowledge, the aim of (Kantian) philosophy is to work out what makes knowledge possible. With Kant, philosophy is no longer dealing directly with the way reality is in itself, but it is restricted to the way reality appears to us through the categories of our understanding. The implication is that philosophy deals only with phenomena, not with the way things are in themselves. This structure that structures phenomena is what Kant called "the transcendental". In this shift, Kant turned philosophical reflection onto itself, and it is because of

this reflexive gesture that it is not enough to assert the results from empirical science as a solution to this problem. This is what Helmholtz did. Even though he was deeply appreciative of Kant, Helmholtz asserted the primacy of 'physiological investigations on sense perception' over and above Kantian philosophy (Helmholtz, 1995: 364). However, the problem formulated by Kant is not a lack of (scientific) knowledge, but the *possibility and the categories of knowing* itself. The insistence on continued experimentation falls short in this case, because the problem is the faculty that grounds the orientation from which we conduct experiments: 'For the principles in accordance with which we set up experiments must themselves always be derived from the knowledge of nature, hence from theory.' (Kant, 2000: 6, first introduction). Said differently, our experiments too presuppose a theoretical position to depart from, and it is this position that Kant questions.

This does not mean a regression to obscurantist anti-science. Empirical science keeps all of its importance, and we definitely need more 'physiological investigations on sense perception', but Kant's point is that even empirical science does not provide us with direct access to reality. If it did, the history of science would not be riddled with detours, trial and error, dead ends, extended periods of little progress, alternated with unsettling leaps forward. Empirical science is without a doubt the privileged way to gain knowledge about the external world, but the results from our experiments are themselves caught in a loop presupposing some understanding of how our object under study functions to start out from, leaving Kant's problem intact. Instead of resigning ourselves to the easy division between "bad" subjective influences and "good" objective experimental results, the challenge becomes to show, and where needed to criticize and to change, how the inclusion of subjectivity – our theoretical position – is constitutive for the objects we attempt to know as "objective" reality. We place "objective" between quotation marks not because there is no such thing as objective reality distinct from human subjectivity, but because our access to "objective" reality is subjectively mediated by our theoretical orientation, as well as technological apparatuses in our experiments. Therefore the philosophical background of our field is not just of interest for historical and philosophical reasons, but it is a matter of understanding what traditions our field of research is beholden to, so that we may more easily revise or even depart from these traditions if they no longer serve us.

If we turn back to cognitive neuroscience and the FEP, the question becomes: what is the traditional view? If we accept the FEP as a possible paradigm shift, what are we shifting away from? It is beyond the scope of this introduction

to investigate the entire philosophical backdrop of our field (cf. Bennett & Hacker, 2022), but we can pinpoint some points of convergence relevant for the present thesis. One common criticism leveraged against the traditional study of perception is the tendency in our experiments to bombard participants repeatedly and passively with arbitrary stimuli. This practice makes complete sense if we subscribe to a view of brain function from which primacy is afforded to external influences over and above internal states. This is the view of British empiricism. Take for instance John Locke's *An Essay Concerning Human Understanding* from 1690:

> "A man begins to have ideas when he first has sensation. [...] If it shall be demanded then, WHEN a man BEGINS to have any ideas, I think the true answer is,—WHEN HE FIRST HAS ANY SENSATION. For, since there appear not to be any ideas in the mind before the senses have conveyed any in, I conceive that ideas in the understanding are coeval with SENSATION; WHICH IS SUCH AN IMPRESSION OR MOTION MADE IN SOME PART OF THE BODY, AS MAKES IT BE TAKEN NOTICE OF IN THE UNDERSTANDING." (Locke, 1796, Book II, Chapter I, §23)

In the paragraph immediately following Locke concludes: "Thus the first capacity of human intellect is,—that the mind is fitted to receive the impressions made on it". A couple centuries later and our experimental practices to study the brain are still beholden to these coordinates. It is this view the FEP breaks with by not only affording primacy to expectations encoded by internal states, on which external "impressions" impinge, but also by asserting the importance of an organism's capacity to elicit its own sensations through the detour of acting on external states. We find similar importance afforded to this fundamental loop in the work of Hegel, a major successor of Kant. **Chapter 4** explores how the FEP links up with alternative historical and philosophical traditions from Kant to Hegel.

In the remainder of this general introduction, we will explore two possible ways to deploy a departure from the empiricist perspective in the subfields revolving around the role of action in perception (§1.3) and conscious perception (§1.4).

1.3. Action in perception and belief updating

In his 2019 book *The Brain from Inside Out*, neuroscientist Georgi Buszáki takes issue precisely with the way this empiricist tradition predominates in present-day neuroscience and psychiatry (Buzsáki, 2019):

> "In the empiricism-inspired model, signals enter the brain from the outside, neuronal circuits process and perceive them, and some part of the brain decides whether or not to generate a motor response. The main emphasis is on perceptual processing and association, which are believed to be the main drivers of the brain's ability to comprehend and represent the outside world." (p. 8)

Buzsaki's core argument is that instead, "the brain is a self-organized system with preexisting connectivity and dynamics whose main job is to generate actions and to examine and predict the consequences of those actions." (p. xiii). This view is entirely in line with the FEP (cf. Friston & Buzsáki, 2016): the shift towards an emphasis on internally maintained and acted on Bayesian models is in essence an inside-out perspective already.

Instead of perception in isolation, action gains particular importance from this perspective because it is the main way for an organism to put its models to the test; to check whether they are still up-to-date, and to change them where necessary. **Chapter 5** is a preliminary attempt at contrasting an inside-out perspective with the traditional outside-in perspective by contrasting conditions in which the prediction of stimuli was either self-generated through action or externally induced by sensory cue, to see how prediction violation and belief-updating differ within the context of a Bayesian model derived from the FEP.

1.4. Conscious perception

Most of conscious perception research has converged on a cortex-centered approach, epitomized in the title of an article by Boly and colleagues from 2017: "Are the neural correlates of consciousness in the front or in the back of the cerebral cortex?". To pose the problem of consciousness in these terms excludes a role for the rest of the brain. The FEP comes with a revaluation of internal states in contrast to the traditional focus on external influences. When we speak of internal states in terms of an organism's generative model, it is important to emphasize that we speak of the entire brain, not just the cortex. The FEP

ascribes particular importance to the basal ganglia (BG) and its irrigation by dopamine for active inference (FitzGerald et al., 2015; Friston et al., 2014; Friston et al., 2012). Friston (2018) has postulated that any organism that engages in the minimization of expected surprise over a sufficiently large timescale, i.e., in selecting between models that predict that what has not yet happened (i.e., as in active inference), exhibits (self-)consciousness.

The structures comprising the BG are well-suited to be investigated for this model-selection purpose, because they not only form a central hub in circuits spanning from brainstem to cortex, but they have long been known to play a role in sensory and perceptual processes (Afrasiabi et al., 2021; Alexander & Crutcher, 1990; Arsalidou et al., 2013; L. L. Brown et al., 1997; Seger, 2013). Indeed, selection has been hypothesized to be the BG's main function (Redgrave et al., 1999, 2011) The idea is then that striatal dopamine influences the contents of conscious perception by determining which 'internal model' or interpretation of the current sensory state dominates perceptual experience, in anticipation of what is to come. **Chapters 2 and 3** comprise two attempts at including subcortical structures into the study of conscious perception.

Chapter 2

Conscious perception and the modulatory role of dopamine: no effect of the dopamine D2 agonist cabergoline on visual masking, the attentional blink, and probabilistic discrimination

---

Published as:

Boonstra, E. A., van Schouwenburg, M. R., Seth, A. K., Bauer, M., Zantvoord, J. B., Kemper, E. M., Lansink, C. S., & Slagter, H. A. (2020). Conscious perception and the modulatory role of dopamine: No effect of the dopamine D2 agonist cabergoline on visual masking, the attentional blink, and probabilistic discrimination. *Psychopharmacology*, *237*(9), 2855–2872. https://doi.org/10.1007/s00213-020-05579-9

2.1. Abstract

Rationale. Conscious perception is thought to depend on global amplification of sensory input. In recent years, striatal dopamine has been proposed to be involved in gating information and conscious access, due to its modulatory influence on thalamocortical connectivity.

Objectives. Since much of the evidence that implicates striatal dopamine is correlational, we conducted a double-blind crossover pharmacological study in which we administered cabergoline – a dopamine D2 agonist – and placebo to 30 healthy participants. Under both conditions, we subjected participants to several well-established experimental conscious-perception paradigms, such as backward masking and the attentional blink task.

Results. We found no evidence in support of an effect of cabergoline on conscious perception: key behavioral and event-related potential (ERP) findings associated with each of these tasks were unaffected by cabergoline.

Conclusions. Our results cast doubt on a causal role for dopamine in visual perception. It remains an open possibility that dopamine has causal effects in other tasks, perhaps where perceptual uncertainty is more prominent.

2.2. Introduction

The relationship between consciousness and the brain is often lauded as one of the big mysteries in contemporary science. How does the brain constrain its own spontaneous activity as well as the influences it undergoes from outside, in the determination of conscious awareness? Several influential theories propose that consciousness is related to the 'broadcasting' of sensory information to the whole brain and that thalamocortical circuits serve as an important mediator of such broadcasting (Crick & Koch, 2003; Dehaene & Changeux, 2011; Edelman, 2003). The broadcasting of sensory information necessitates the occurrence of selection or filtering, simply because not everything which takes place in the brain reaches conscious awareness.

With the requirement of selection, we cannot ignore the role of the basal ganglia: a cluster of subcortical nuclei located deep in the brain, which modulate activity of thalamocortical circuits (Smith, Raju, Pare, & Sidibe, 2004) and have long been implicated in action selection; the decision to execute one of several possible behaviors (Redgrave, Prescott, & Gurney, 1999). Notably, the basal ganglia are connected through parallel loops via the thalamus not only to the motor cortex, but to many parts of frontal cortex as well (Alexander, DeLong, & Strick, 1986). Hence, it is capable of modulating a wide range of cognitive operations. Indeed, the basal ganglia have been implicated in working memory updating (Frank & O'Reilly, 2006), attention shifting (Cools, 2011), and visual categorization (Seger, 2008): cognitive acts that support suggestions concerning the common principle underlying the basal ganglia's operations; namely, selection (Frank, Loughry, & O'Reilly, 2001; Redgrave, Prescott, & Gurney, 1999).

The striatum – the biggest structure constituting the basal ganglia – deserves special attention on this topic. As the basal ganglia's primary input nucleus, the striatum is well-positioned to play a pivotal role in the basal ganglia's selective functionalities, as terminal fields from different cortical regions converge in the striatum (Yeterian & van Hoesen, 1978; Haber, Kim, Mailly, & Calzavara, 2006; Mailly, Aliane, Groenewegen, Haber, & Deniau, 2013; Heilbronner, Meyer, Choi, & Haber, 2018). While dopaminergic projections are usually associated with reward prediction error (Schultz, 2016), and the role of reward in perceptual decision making (Ding & Gold, 2013); less well-studied signaling of the (dopamine-infused) striatum include saliency, threat, processing of sensory information, and promoting of behaviors reliant on sensory information (Cox & Witten, 2019).

Despite the suggestion that the basal ganglia may be "mute" as regards to consciousness (Boly et al., 2017), it has been known for a long time that basal

ganglia structures are involved in sensory and perceptual processes (Alexander & Crutcher, 1990; Arsalidou, Duerden, & Taylor, 2013; L. L. Brown, Schneider, & Lidsky, 1997; Seger, 2013). Both the striatum and dopaminergic firing have been implicated in the tight relationship between perception and action, and consciousness more generally. For example, dopamine-depleted mice are unable to attend to salient sensory information and choose appropriate actions, suggestive of a critical role for dopamine in the expression of consciousness (Palmiter, 2011). In humans as well, it has recently been shown that minimally conscious patients suffer from a dopaminergic deficit in presynaptic neurons projecting to the striatum and central thalamus (Fridman, Osborne, Mozley, Victor, & Schiff, 2019).

In humans, the striatum and its irrigation by dopamine have also been implicated in well-known experimental paradigms used to study the neural correlates of consciousness, such as the attentional blink and backward masking task. In the attentional blink task, a deficit occurs when people have to detect two target stimuli (T1 and T2) presented in close temporal succession among distracter events. Specifically, when T2 follows T1 within 100–500 ms, it often goes unnoticed. This deficit is called the attentional blink (AB; Shapiro, Raymond, & Arnell, 1997). Healthy participants with more D2-like receptor binding in the striatum – as shown with PET – showcased a larger AB (Slagter et al., 2012). In addition, intracranial EEG recordings in the ventral striatum revealed a short-latency increase in theta-band oscillatory activity only for consciously perceived target stimuli (Slagter et al., 2017).

In backward masking tasks, processing of target stimuli is interrupted by presenting a mask in close succession to the target (Breitmeyer, 2007). Studies employing fMRI consistently show differences in BOLD activity in the striatum and thalamus between seen and unseen stimuli using backward masking tasks (Bisenius, Trapp, Neumann, & Schroeter, 2015). In another PET study, dopamine D2 binding potential in the right striatum was found to correlate positively with both objective (task performance) and subjective (seen/unseen) visibility during backward masking (Van Opstal et al., 2014). These studies collectively suggest a role for the striatum and dopaminergic activity in the selection of visual information and the formation of conscious visual percepts.

However, up until now the relationship between striatal dopamine and conscious perception is based on correlational evidence. As such, in the present study we sought to manipulate this relationship experimentally, by administering the dopamine D2 agonist cabergoline to healthy participants. Out of the two main dopamine receptor families, D2 receptors have been found to be more prevalent in the striatum, while D1 receptors are present

more in the prefrontal cortex (Gerfen, 1992). We chose cabergoline because it has greater affinity for D2 receptors, and it has been reported to have less side-effects compared to other D2 agonists such as bromocriptine (Frank & O'Reilly, 2006). Cabergoline, at low doses, has been suggested to preferentially stimulate pre-synaptic D2 autoreceptors, which have been found to inhibit phasic dopamine bursts in the striatum (Ford, 2014; Frank & O'Reilly, 2006). Cabergoline has been successfully administered in small dosages (1-1.5 mg) to manipulate performance in healthy participants on working memory tasks (Broadway, Frank, & Cavanagh, 2018; Fallon, Zokaei, Norbury, Manohar, & Husain, 2017; Frank & O'Reilly, 2006), a modified version of the Simon task (Cavanagh, Masters, Bath, & Frank, 2014), as well as action cancellation and error awareness tasks (Nandam et al., 2013). When cabergoline was administered to Parkinson's patients for longer periods of time, decreases contrast sensitivity was found (Hutton, Morris, & Elias, 1999). We administered cabergoline in an attempt to manipulate performance on two paradigms traditionally used to study the neural correlates of conscious perception: backward masking and the attentional blink task. In addition, there is increasing evidence that dopamine plays a crucial role in determining the influence of sensory information in relation to expectations acquired through past experience (Cassidy et al., 2018; Friston et al., 2012). To investigate this relationship, we subjected participants to a probabilistic discrimination task in which the probability of stimulus occurrence varied across blocks of trials, thereby tapping into the learning capabilities the basal ganglia is implicated in traditionally (Berke, 2018).

The difficulty with manipulating dopamine is that effects have been found to depend on baseline dopamine levels in accordance with an inverted-U-shape (Cools & D'Esposito, 2011). This means that dopamine is thought to have an optimal level for task performance, but that dopamine manipulations may either benefit or worsen performance dependent on an individual's starting point on the U-curve. Two measures used to estimate baseline dopamine levels are working memory operation span (OSPAN; Broadway et al., 2018; Cools, Gibbs, Miyakawa, Jagust, & D'Esposito, 2008) and spontaneous eye-blink rate (sEBR; Cavanagh et al., 2014; Jongkees & Colzato, 2016). We used these measures to analyze and control for potentially different effects of cabergoline in relation to baseline dopamine levels.

2.3. Methods

Participants

30 native Dutch-speakers were recruited from the University of Amsterdam subject pool to complete the experiment (mean age = 22, range 18-29, 25 female). Because our study is the first to investigate the effects of cabergoline on conscious perception, we did not conduct a power analysis. Instead, our sample size is on the upper end of sample sizes employed by previous research in which cognitive and neural effects of cabergoline were reported with 12-30 participants (Cavanagh et al., 2014; Cohen, Krohn-Grimberghe, Elger, & Weber, 2007; Frank & O'Reilly, 2006; Nandam et al., 2013; Norbury, Manohar, Rogers, & Husain, 2013; Yousif et al., 2016; Fallon et al., 2017; Broadway et al., 2018), and in which the behavioral and event-related potential (ERP) effects were reported which we aimed to manipulate with cabergoline (12-15 subjects; Del Cul et al., 2007; van Opstal et al., 2014; Slagter et al., 2012).

Four participants experienced adverse reactions during the cabergoline session (dizziness and nausea) that interfered with their ability to participate in the study. In two participants, nausea was present to the point of vomiting. One participant dropped out due to headaches in the placebo session. Completers and non-completers did not differ in baseline age, BMI, heart rate/blood pressure, or baseline dopamine proxies (see below). It should be noted however that all drop-outs were women.

In total, 124 individuals were considered for inclusion in the study. 113 were interviewed over the telephone to ensure normal or corrected-to-normal vision, no history of neurological, psychiatric, or any other relevant medical problems, and abstinence from psychoactive medication. The use of hormonal contraceptives served as an additional inclusion criterium for female participants, who were tested outside of their period due to variability in D2R availability during different phases of the menstrual cycle (Czoty et al., 2009). 63 individuals were eligible for participation, out of which 35 participants were available to schedule the required three lab visits (see Procedure). 5 participants did not meet requirements for inclusion based on the first lab visit. After excluding 5 participants with adverse reactions, the final sample consisted of 25 participants (mean age = 22, range 18-29, 20 female). In exchange for their participation, participants received 10 euros an hour, with a minimum of 110 euros in total. The study protocol was approved by the medical ethical committee of the Academic Medical Centre, Amsterdam (currently Amsterdam University Medical Centers). All participants provided written informed consent in accordance with the Declaration of Helsinki.

Procedure

Participants came to the lab three times on different days (see Table 1). Once for screening (duration: 2.5h), and twice for an experimental session in which either placebo or 1.5mg of cabergoline was administered orally (duration: 4.5h each), as part of a double-blind crossover design. Previous studies found cognitive and neural effects of cabergoline using a dosage of 1-1.5mg (Cavanagh et al., 2014; Cohen et al., 2007; Frank & O'Reilly, 2006; Nandam et al., 2013; Norbury et al., 2013; Yousif et al., 2016; Fallon et al., 2017; Broadway et al., 2018). As with our sample size, we chose a dosage of 1.5mg to be on the upper end of previously employed dosages. There was at least a day in between screening and the first session, and at least a week between both sessions.

Screening. The first lab visit took place anywhere between 09:00 and 17:30. After providing written informed consent, participants answered a series of questions concerning potential medical conditions. Next, we conducted the M.I.N.I.; a structured screening interview for DSM-IV axis-I disorders (Sheehan et al., 1998). We subsequently measured participant's weight, height, BMI, blood pressure (BP), and heart rate. Participants were included in the experiment only if these measures fell within pre-established bounds (BMI 18-30, diastolic BP < 50 or > 90 mmHg, systolic BP < 95 or > 140 mmHg). Next, six external electrodes were attached to the participant's face and ears in order to measure spontaneous eye-blink rate (sEBR) at rest. Finally, participants completed an operation span (OSPAN) working-memory task (Unsworth, Heitz, Schrock, & Engle, 2005), and a titration procedure for two behavioral tasks to be completed during both experimental sessions (backward masking and probabilistic discrimination; see below).

Session. All placebo and cabergoline sessions (the second and third visit) took place between 08:30 and 14:00. Participants were instructed to abstain from drug and heavy alcohol use, the day before and during the day of the session. Also, participants were instructed to abstain from caffeine and nicotine the morning of the session. Compliance to the instructions was checked by the examiner on arrival, in case of non-compliance the session was postponed. Female participants completed a midstream pregnancy test. Breakfast was offered, in order to avoid cabergoline-intake on an empty stomach. Blood pressure and heart-rate were measured using an Omron® M3 comfort Sphygmomanometer, and participants filled in a visual analogue scale (VAS, see below) three times during the session: on arrival, at around 1.5h after placebo or cabergoline intake, and at the end of the session. After the initial blood pressure/heart-rate/VAS measurement, participants were administered either placebo or cabergoline in a double-blind fashion (order randomized

across participants). After a 40-minute break, a BioSemi ActiveTwo system (BioSemi Inc., Amsterdam, The Netherlands) EEG cap and electrodes were fitted. Drug plasma levels have been found to reach maximum concentration after approximately 1.5-3h (Persiani, Rocchetti, Pacciarini, Holt, Toon, & Strolin-Benedetti, 1996; Agúndez, Garcia-Martin, Alonso-Navarro, & Jiménez-Jiménez, 2013). Approximately 1 hour and 20 minutes after drug intake, participants completed 6 minutes of sEBR recordings, followed by the backward masking task around 1.5h after drug intake (see below), during which EEG was recorded. After this task, the EEG setup was removed from the participant's head. After a 30-minute lunch, participants proceeded to the attentional blink task, a simple reaction time task, and the probabilistic discrimination task (see below). At the end of the experiment, one final blood pressure/heart rate and VAS measure was undertaken. At the end of the final session, participants were asked to indicate in which session they believed they had received cabergoline.

Table 1. *Experimental procedures for the screening, placebo, and cabergoline sessions.*

Screening	**Sessions (placebo/cabergoline)**
Medical questionnaire	Questions regarding recent substance use + pregnancy test
M.I.N.I.	Blood pressure/heart rate + VAS 1
Body weight/height	Drug intake
Blood pressure/heart rate	40-minute break
sEBR	EEG setup
OSPAN	sEBR
Titration backward masking	Blood pressure/heart rate + VAS 2
Titration probabilistic discrimination	Backward masking (± 1.5h after drug intake)
	30-minute lunch
	Attentional blink
	Reaction time task
	Probabilistic discrimination
	Blood pressure/heart rate + VAS 3 (± 3.5h after drug intake)

Physiological and subjective state measures

Heart rate and blood pressure. Physiological measurements were taken once during screening, and three times during both sessions; namely, on arrival, at around 1.5h after drug intake, and on completion of testing (± 3.5h after drug intake) (see Table 1). These measurements were obtained using an Omron® M3 comfort Sphygmomanometer.

Subjective self-report. A set of sixteen VAS measures were used (Bond & Lader, 1974), to assess the subjective state of the subject before medication intake, at around 1.5h after drug intake, and on completion of testing (± 3.5h after drug intake). Each scale consisted of a 100-mm horizontal line, anchored by contrasting states of mind (e.g., happy versus sad). Subjects were asked to regard each line as a continuum and to rate their feelings at the time by moving a vertical slider across each line. The scales could then be scored by measuring the length in millimeters from the positive end of each line to the subject's marked location. These sixteen VAS measures were summarized as three categories: contentedness, calmness, and alertness (Bond & Lader, 1974).

Baseline dopamine proxies. Both sEBR and OSPAN are widely-used measures that have been related to baseline dopamine levels (Cools & D'Esposito, 2011; Jongkees & Colzato, 2016). Both measures have been used in combination with cabergoline in order to account for individual differences in baseline dopamine (Broadway et al., 2018; Cavanagh et al., 2014). Eye blink rate is defined as the number of spontaneous eye blinks per minute. The measure has high test-retest reliability (Kruis, Slagter, Bachhuber, Davidson, & Lutz, 2016) and is an often-used biomarker of baseline dopamine D2 receptor functioning (Jongkees & Colzato, 2016; Karson, 1983; Taylor et al., 1999, but see Sescousse et al., 2018). Subjects were asked to look at a central fixation cross on a computer screen in a relaxed state for 6 minutes while we measured eye activity from a set of vertical and horizontal electrodes, in order to detect eye blinks. This procedure was employed during all three lab visits.

Eye blinks were established in two ways. First, through a fully automatic procedure implemented in the python module MNE (*create_eog_epochs*; Gramfort et al., 2013). Second, eye-blinks were established through a custom semi-automatic procedure using EEGLAB for MATLAB (Delorme & Makeig, 2004). If mean sEBR per minute differed more than 3 blinks between both methods, the semi-automatic procedure was repeated, and an average was taken of both semi-automatic attempts as the final value. Prior to any repetition of the semi-automatic method, correlations between the automatic and semi-automatic method exceeded .95 for measurements during all three lab visits.

OSPAN is a working-memory task with high test-retest reliability (Unsworth et al., 2005), in which participants are instructed to remember letters, while solving simple arithmetic problems in between letter presentation (Unsworth et al., 2005). Sets of 3-7 letters were presented successively at fixation. The OSPAN score was calculated through partial credit scoring, so that each correctly recalled letter in the appropriate location was counted as correct, regardless of whether the entire sequence was recalled correctly or not. Scores could range from zero to 75. OSPAN was measured only during screening.

Reaction time. To assess the effects of cabergoline on alertness we administered a 40-trial simple reaction time (RT) task (Brown et al., 2016). In this task, participants had to respond as quickly as possible by pressing the spacebar whenever a white circle (subtending approximately 3.1° of visual angle) appeared at the center of the computer screen against a black background. Stimulus onset asynchrony was jittered between 500 and 1250 ms, with a mean of 1000 ms. This task lasted less than 2 minutes.

Main experimental paradigms

All stimuli were presented on an ASUS VG236H 23-inch LCD screen (refresh rate = 100 Hz, resolution 1920x1080). Participants viewed the screen at a distance of 80 cm.

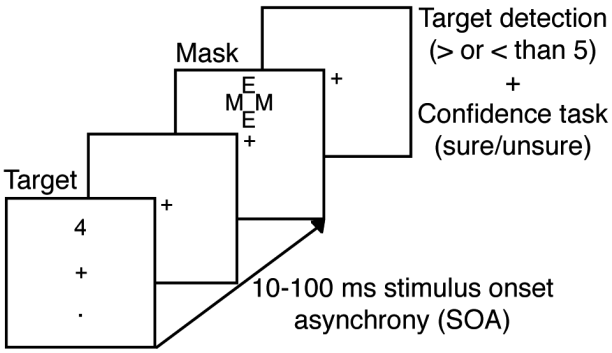


Figure 1. Experimental procedure for one trial of the backward masking task (adapted from Van Opstal et al., 2014)

Backward masking. In the backward masking task, adapted from Van Opstal et al. (2014), participants had to indicate whether briefly presented masked digits (1, 4, 6, or 9) were smaller or larger than 5 and rate the confidence in their response (Figure 1). Each trial started with the presentation of a central fixation cross (30 point Courier New), which increased in size (106 point Courier

New, 150 ms duration), cueing the impending target. The target stimulus (30 point Courier New) then appeared for 10 ms at one of two positions centered at the vertical midline (top or bottom, 2.29° from fixation). Both stimulus locations were equally probable. A mask followed the target (200 ms duration) at a variable stimulus onset asynchrony (SOA). Due to the employed refresh rate of 100Hz, the SOA could vary from 10 ms to 100 ms in 10 ms steps. By making the delay between cue and target dependent on SOA, the delay between cue and mask was held constant at 800 ms. The mask (30 point Courier New) was composed of two letters "E" and two letters "M", tightly surrounding the target location without superimposing or touching it. All stimuli were black and presented on a white background, using the Psychophysics toolbox for MATLAB (Brainard, 1997). The central fixation cross was visible throughout the experiment.

Participants were instructed to indicate by button press whether the presented digit was smaller or larger than 5, while simultaneously indicating the confidence in their response (sure/unsure); resulting in four possible responses (<5 sure, <5 unsure, >5 unsure, >5 sure). In previous research it was found that the D2 agonist pergolide affected response confidence (Lou et al., 2011). Responses were given by means of a response box attached to the arm rests of the participant's chair. Response buttons were counterbalanced across participants, who were instructed to guess one of two "unsure" buttons if they did not see the target.

If the participant's reaction time exceeded 1 second, a message was presented indicating that their response was too slow for the duration of 1 second, urging a faster response. An individual threshold for awareness was established during the screening session (see above), by fitting a logistic model (threshold defined as SOA corresponding to 75% accuracy; mean threshold = 52.93 ms, min = 31 ms, max = 89 ms, sd = 14.38; Del Cul, Dehaene, & Leboyer, 2006; Van Opstal et al., 2014). This model was fitted on the basis of 176 trials during screening, where each of 11 SOA durations (from 0-100 ms) was presented 16 times.

Prior to the experiment, participants first completed a practice block (176 trials in screening, 88 trials during placebo and cabergoline sessions). In both drug sessions, participants completed 920 trials in total, split by seven possible SOAs between target and mask: 200 mask-only trials (0 ms SOA), 200 trials each for the main SOAs (10 ms / awareness threshold / 100 ms), 40 trials surrounding the threshold (threshold minus 10 ms and threshold plus 10 ms), and 40 trials with a 70 ms SOA. The trials in between individual thresholds and 100 ms were excluded from analysis (threshold + 10ms and 70 ms), because some participants arrived at an individual threshold at or above 70 ms. This meant that trials with

a SOA at threshold + 10 ms and 70 ms would fall below or above participant's individual threshold, depending on the participant. 80 trials per participant were discarded for this reason, leaving 840 trials.

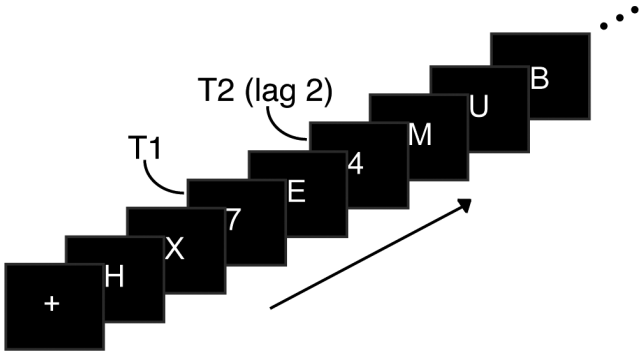


Figure 2. Experimental procedure for one trial of the attentional blink task (adapted from Slagter et al., 2017)

Attentional blink. Participants also performed a standard AB task in which they had to identify two digits (T1 and T2) presented in a rapid stream of centrally presented distractors (letters and symbols; adapted from Slagter et al., 2012; Slagter et al., 2017; see Figure 2). T2 followed T1 either in the time window of the AB, after 200 ms (short-interval trial), or outside the time window of the AB, after 800 ms (long-interval trial). Each trial started with a central fixation cross (1500 ms), after which the stimulus stream began, consisting of 22 stimuli. Stimuli were presented on a black background (RGB 70, 70, 70) at the center of the screen (28 point Arial; 0.85° visual angle) for 50 ms, followed by a 50 ms blank. Digits were drawn randomly (without replacement) from the set 2–9. Distractors were randomly drawn (without replacement) from the following set of 30 letters and symbols: W, E, R, T, Y, U, P, A, D, F, G, H, J, K, L, Z, X, C, V, B, N, M, @, #, $, %, }, &, <, and =. Participants were asked to indicate sequentially the identity of the targets they saw, using the numpad on a standard keyboard. If they missed a target, they were instructed to guess. Stimulus presentation was performed using Presentation (Neurobehavioural Systems).

In both sessions, participants first completed a short practice block (20 trials), in which the first 8 trials moved at half speed. Next, participants moved on to the main experiment (222 trials), spread over 6 blocks consisting of 37 trials each.

Probabilistic discrimination. In the probabilistic discrimination task, adapted from Bauer et al. (2016, September), participants were presented

continually with a central fixation cross (28 point Arial, RGB 0, 0, 0), on top of which an image (6.68° visual angle) of either a face or house was presented for 120 ms, against a grey background (RGB 128, 128, 128). Face stimuli were created on the basis of the Park Aging Mind Laboratory, University of Texas at Dallas (Minear & Park, 2004), while house stimuli were based on the Caltech University Computational Vision database (http://vision.caltech.edu/archive.html). On each trial, participants had to report the category of the image with by pressing "Q" or "P" on a standard keyboard. Responses were counterbalanced across participants, who were instructed to emphasize accuracy over and above speed. The maximum response interval was 1700 ms, after which the next stimulus was presented regardless of whether a response was given. The inter-trial interval was jittered and varied from 800 to 1200 ms.

The difficulty of stimulus discrimination was manipulated in terms of stimulus coherence (Figure 3). Stimuli from the above-mentioned databases were cropped to the outlines of faces and houses and a 2-dimensional spatial Fourier transform (on luminance values for x-/y-coordinates) was calculated. The amplitude (power-) spectra of all 82 face and house images were averaged and subsequently applied to all individual images, such that all images (faces and houses of all coherence levels) had an identical power spectrum. In other words, none of them differed in global contrast or luminance. The phase-spectra of each individual image (that therefore provided all pictorial information) was retained and was subsequently superimposed with various levels of (uniform) random noise for each image (Bauer et al., 2016, September). To account for bias in the circular phase-distribution of superimposed noise and signal phase-spectra, noise-spectra were sampled following previous suggestions (Dakin, Hess, Ledgeway, & Achtman, 2002).

Titration consisted of two subsequent procedures, in order to establish three difficulty levels for each individual participant. First, participants completed 300 trials, spread over 10 staircase blocks, in order to establish difficulty levels corresponding to an accuracy of 75% for each stimulus category separately. During this 3-up-1-down staircase procedure, participants received a green thumbs-up (RGB R [56 154 79], border RGB [17 79 22]) or red thumbs-down (RGB [83 2 5], border RGB [251 84 84]) at the center of the screen as feedback after each trial (500 ms duration).

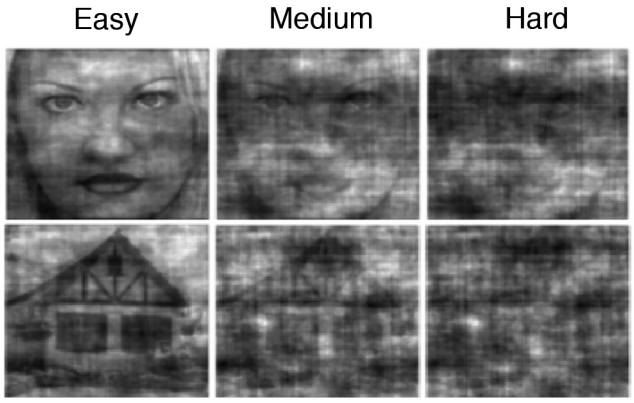


Figure 3. Probabilistic discrimination stimuli (adapted from Bauer et al., 2016, September)

Next, a total of 810 trials followed, across 27 blocks, in order to extrapolate the acquired difficulty level to three difficulty levels. For the second part of the titration procedure, participants received feedback in the break in between blocks; in order to counteract the development of a bias for one of two response categories. Difficulty levels were estimated using the method of constant stimuli (MOCS; Bauer et al., 2016, September). The psychometric functions obtained through this procedure were used to estimate difficulty (coherence) levels corresponding to 70, 82, and 95% accuracy.

In both the placebo and cabergoline session, the same three difficulty levels were employed that were acquired from the titration procedure during screening. Unbeknownst to participants, the prior probability of each category was manipulated in a block-wise manner (20/35/50/65/80%), spread over 25 blocks of 40 trials each, for a total of 1000 trials per session. As such, we manipulated perceptual information (difficulty) and stimulus prior probability independently of one another. Stimulus presentation for the titration procedure during the screening was performed using the Psychophysics toolbox for MATLAB (Brainard, 1997), and Presentation was used to present the task in both experimental sessions (Neurobehavioural Systems).

Behavioral analyses

Physiological and subjective state measures. In order to test whether physiology and subjective state changed over the course of the experiment, and whether these measures were influenced by cabergoline, we conducted a repeated-measures analysis-of-variance (RM ANOVA) for heart rate, diastolic and systolic blood pressure, and each VAS category separately; across all three time-points and both sessions. We conducted a paired-samples t-test between

sessions for our simple RT task, as an additional measure to assess alertness. In order to test whether cabergoline exerted influence on sEBR, we conducted a paired-samples t-test between sEBR under placebo versus cabergoline. Kendall's Tau correlation was employed to establish the relationship between sEBR sessions, as well as the relationship between sEBR and OSPAN, as this coefficient is more robust in the case of small samples and tied ranks (Bonett & Wright, 2000). Correlations were Bonferroni corrected for multiple comparisons.

Backward masking. The dependent measures in the backward masking task were accuracy (0/1) and confidence (unsure/sure). For both of these measures, we computed a 2 (Drug; placebo/cabergoline) x 5 (SOA; mask-only/10 ms/threshold – 10 ms/threshold/100 ms) RM ANOVA. Furthermore, we performed an additional analysis including screening sEBR and OSPAN as covariates in both of these analyses, as we predicted cabergoline effects may depend on individual baseline dopamine levels, based on previous reports (Broadway et al., 2018; Cavanagh et al., 2014; Cools & D'Esposito, 2011; Jongkees & Colzato, 2016). We repeated these analyses for the two-alternative forced choice version of the signal detection theory parameters d' (Green & Swets, 1966) and meta-d' (Maniscalco & Lau, 2012), for the three primary SOAs (10 ms, threshold, and 100 ms).

One participant confused the confidence response buttons in one session. We reversed these confidence scores manually. One participant experienced side effects only near the end of the last session. This participant thus completed the backward masking task twice without knowledge about drug condition. In order to maximize statistical power, this participant was included in all analyses concerning the backward masking task (including EEG), but not in the analyses of other tasks. It did not matter for our results whether this participant was included in the backward masking task analyses or not.

Attentional Blink. The dependent measures for the AB task were T1 accuracy and T2 | T1 accuracy. In other words, T2 accuracy was based only on those trials where T1 was correctly reported. For each of these measures, we computed a 2 (Drug; placebo/cabergoline) x 2 (Lag; 2/8) RM ANOVA. For this task as well, we computed additional analyses in order to include sEBR and OSPAN as covariates. Finally, we computed AB size, in order to investigate the relationship between AB size and our baseline dopamine measures as some (Colzato, Slagter, Spapé, & Hommel, 2008) but not other studies (Slagter et al., 2012) have found.

Probabilistic discrimination. In the case of the probabilistic discrimination task, our dependent measure of interest was accuracy. As such, we computed a 2 (Drug; placebo/cabergoline) x 3 (Difficulty; easy/medium/hard) x 5 (Probability;

.2/.35/.5/.65/.8) RM ANOVA. For this analysis as well, we computed additional models including screening sEBR and OSPAN as covariates.

Our titration procedure was not successful for all participants. As a result, a number of participants ended up with only two difficulty levels for one out of two stimuli. For this reason, we repeated the above analysis for both the group with all difficulty levels (N = 16), and participants who had either two or three difficulty levels as a result of the titration procedure (N = 24). In this latter analysis, we excluded the medium difficulty trials. One participant was excluded from both analyses, because this participant ended up with only one difficulty level for face stimuli.

All covariates in the above-mentioned RM ANOVAs were centered (van Breukelen & van Dijk, 2007). For all repeated-measures ANOVA analyses, whenever Mauchly's test suggested a violation of sphericity, we report Geenhouse–Geisser corrected *P*-values, but uncorrected degrees of freedom. In order to test for order effects, we repeated each of the RM ANOVAs for our behavioral paradigms including a between-subject factor indicating whether a participant received either placebo or cabergoline in the first session.

Bayesian statistics. In order to evaluate evidence in favor of our (null) hypotheses, we conducted Bayesian statistics. For each reported frequentist test, we report the Bayes factor corresponding to the inclusion of a factor or interaction within the model in question (shortened to BF_{incl}), compared to equivalent models stripped of the effect. For example, $BF_{incl} = 10$ indicates that a model including the factor in question is ten times more likely given the data compared to a model without the variable. Conversely, $BF_{incl} = .1$ indicates that a model without said effect is ten times more likely given the data. All Bayesian statistics were conducted using JASP (2019, version 0.10.0).

All data visualization was performed with the help of raincloud plots (Allen, Poggiali, Whitaker, Marshall, & Kievit, 2019), which include the mean, individual data points, as well as the overall distribution of the measure in question.

EEG

Recording and preprocessing. EEG data, digitized at 512 Hz, were continuously recorded in both the placebo and cabergoline session during 6 minutes of sEBR and the backward masking task, using an ActiveTwo system (BioSemi, Amsterdam, the Netherlands), from 64 scalp electrodes placed according to the 10/20 system, four electro-oculographic electrodes placed above and below, and to the side of the eyes, and two external electrodes attached to each earlobe. EEG data were offline referenced to the average activity recorded

at the earlobes, and high-pass 'firws' filtered (default settings) at 0.05 Hz using a Kaiser window, following previous suggestions (Widmann, Schröger, & Maess, 2015). The continuous data were subsequently epoched from -1.5 to 1.5 s around stimulus presentation and baseline corrected to the average activity between -200 ms and 0 ms pre-stimulus. Epochs containing EMG artifacts or eye blinks surrounding stimulus presentation were rejected based on visual inspection. Extremely noisy or broken channels were interpolated. Remaining eye blink artifacts were removed by decomposing the EEG data into independent sources of brain activity using an Independent Component Analysis, and removing eye blink components from the data for each subject individually. Epochs were low-pass filtered at 30 Hz for visualization purposes only. Preprocessing was done using the Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) for Matlab (The MathWorks, Inc. Natick, MA, USA) using custom-written Matlab scripts.

Analyses. To determine the effect of our manipulations on ERP markers of information-processing, we examined the effects of SOA and Drug on the amplitude of the visual-evoked P1 and N1 components, the N2, as well as of the later P3b (Del Cul, Baillet, & Dehaene, 2007). In line with Del Cul et al (2007), we epoched the ERP data to the onset of the mask. Next, we subtracted the data from the mask-only SOA condition from all other SOA conditions. Finally, we shifted ERP onset back to target onset, in order to compute target-locked ERPs. Visual inspection of the grand- and condition-average ERPs showed that the P1 and N1 components peaked over lateral occipitoparietal scalp sites (PO7, PO3, O1, PO4, PO8, O2), the N2 over centroparietal scalp regions (C1, Cz, C2, CP1, CPz, CP2), and the P3b over central parietal scalp sites (P1, Pz, P2, PO3, POz, PO4). These scalp sites were used to determine the peak amplitude and latency of these components for each condition of interest. Specifically, the largest positive voltage value between 75–150 ms post-target, and the largest voltage negativity within 150–225 ms were selected to determine the amplitude and latency of the P1 and N1 peaks, respectively, for each subject separately. In the case of the P2 and N2, these intervals were 175-250 ms, and 250-375 ms, respectively. For the P3, an interval of 300-450 ms was used. All average amplitude values 15 ms around the peak sample, as well as individual latencies were entered into separate RM ANOVAs with two within-subject factors: SOA (10 ms, threshold-10 ms, threshold, 100 ms), and Drug (placebo/cabergoline). Because the mask-only condition is used to acquire ERP data for the remaining four SOA conditions (see above), the SOA factor contains four instead of five levels for ERP analyses.

2.4. Results

Physiological and subjective state measures

First, we aimed to establish whether cabergoline exerted physiological effects (Figure 4). Heart rate decreased over time in both placebo and cabergoline conditions ($F_{(2, 48)} = 11.8$, $p < .001$, $\eta^2 = .06$, $BF_{incl} > 100$), and was overall higher in the cabergoline condition ($F_{(1, 24)} = 9$, $p = .006$, $\eta^2 = .03$, $BF_{incl} = 57.3$), but there was no interaction between Time and Drug ($F_{(2, 48)} < 1$, $p = .63$, $BF_{incl} = .14$; see Figure 4a). In the case of diastolic blood pressure as well, we also observed a decrease over time ($F_{(2, 48)} = 21.7$, $p < .001$, $\eta^2 = .14$, $BF_{incl} > 100$), but in this case an overall lower measurement in the cabergoline condition ($F_{(1, 24)} = 9.9$, $p = .004$, $\eta^2 = .03$, $BF_{incl} = 12.4$), and again no interaction between Time and Drug ($F_{(2, 48)} < 1$, $p = .42$, $BF_{incl} = .2$). With regards to systolic blood pressure, we also observed a decrease in blood pressure over time ($F_{(2, 48)} = 36.7$, $p < .001$, $\eta^2 = .22$, $BF_{incl} > 100$) and an overall lower measurement in the cabergoline condition ($F_{(1, 24)} = 9$, $p = .006$, $\eta^2 = .03$, $BF_{incl} = 31.2$), but here a significant interaction between Time and Drug was found ($F_{(2, 48)} = 7.2$, $p = .002$, $BF_{incl} = 5.8$; see Figure 4b). Thus, all three physiological measures decreased over time. While heart rate was higher, and diastolic blood pressure lower across the cabergoline session, cabergoline can only be said to have decreased systolic blood pressure, as indicated by the interaction between Time and Drug.

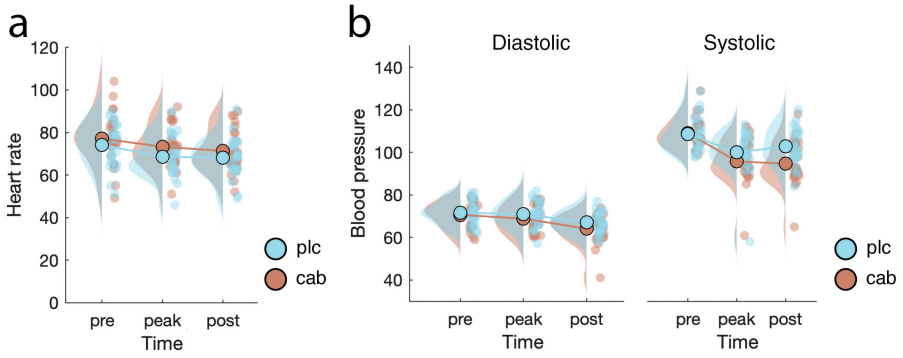


Figure 4. Time course and difference between drug conditions in heart rate and blood pressure. All three physiological measures decreased over time. Cabergoline only affected systolic blood pressure, while heart rate was higher, and diastolic pressure was lower in the cabergoline versus placebo session

Cabergoline also affected the subjective state of participants over the course of the experiment, as indicated by the VAS (visual analog scale, see Figure 5). On the calmness scale, there was no main effect of Drug ($F_{(1, 24)} < 1$, $p = .63$, $BF_{incl} = .23$),

but calmness increased over the course of the experiment ($F_{(2, 48)}$ = 14.5, p < .001, η^2 = .04, BF_{incl} = 31). In addition, we found an interaction between Drug and Time ($F_{(2, 48)}$ = 5.8, p = .005, η^2 = .02, BF_{incl} = 2), reflecting the fact that calmness ratings increased less in the cabergoline condition (Figure 5). In the case of contentedness, there was no main effect of Drug ($F_{(1, 24)}$ < 1, p = .46, BF_{incl} = .3), Time ($F_{(2, 48)}$ = 2.7, p = .08, BF_{incl} = .33), or an interaction between Drug and Time ($F_{(2, 48)}$ < 1, p = .53, BF_{incl} = .16; see Figure 5). Finally, alertness decreased over the course of the experiment ($F_{(2, 48)}$ = 10.3, p = .001, η^2 = .12, BF_{incl} > 100) and was generally lower in the cabergoline session compared to placebo session ($F_{(1, 24)}$ = 6.3, p = .02, η^2 = .01, BF_{incl} = 1.1), but there was no interaction ($F_{(2, 48)}$ = 1.8, p = .18, BF_{incl} = .21; see Figure 5). Thus, cabergoline decreased self-reported ratings of calmness and alertness as the experiment progressed.

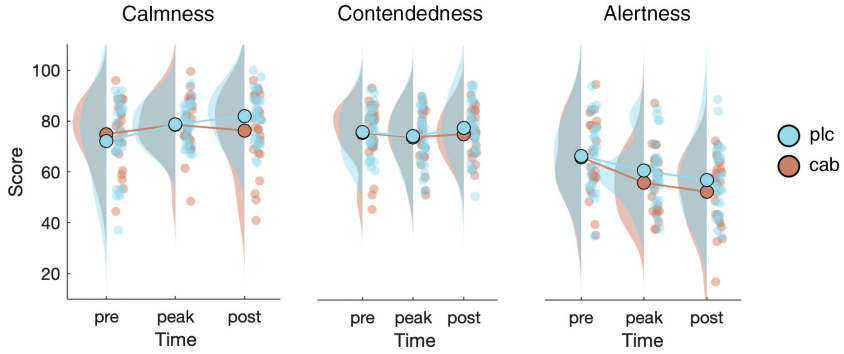


Figure 5. Time course and difference between drug conditions in VAS scores for calmness, contentedness, and alertness. Participants become calmer and less alert over the course of the experiment. Cabergoline stifled this increase in calmness

85% of participants successfully guessed when they received cabergoline at the end of the experiment. However, despite these effects, we found no significant difference in sEBR (spontaneous eye blink rate) between placebo (M = 14 blinks per minute, SD = 8.9) and cabergoline (M = 15.3, SD = 10.6) ($t_{(24)}$ = -1.2, p = .23, BF_{01} = .41). In addition, we were unable to replicate the finding by Cavanagh and colleagues (2014) that baseline sEBR conditioned a cabergoline-induced shift in line with an inverted-U-shape pattern, to the point where high baseline sEBR (as indicative of high tonic striatal dopamine) is associated with a reduction in blink rate, whereas low baseline sEBR is linked to an increase in blink rate: we found that the difference in sEBR between drug conditions was unrelated to screening sEBR (M = 13.2, SD = 7.8) ($\tau_{(25)}$ = .23, p = .11).

Replicating previous findings (Jongkees & Colzato, 2016; Kruis et al., 2016), sEBR was correlated across sessions: between the placebo and cabergoline session ($\tau_{(25)}$ = .63, p < .001), between the screening and cabergoline session ($\tau_{(25)}$ = .37, p = .01), as well as between the screening and placebo session ($\tau_{(25)}$ = .32, p = .03; see Figure 6a). The latter two correlations did not survive a Bonferroni correction for multiple comparisons using a corrected alpha level of .05 / 6 = .0083, based on all correlations computed in this section. Together these results lend support to the robustness of the measure when measured at the same time of day (Barbato et al., 2000; Jongkees & Colzato, 2016). However, despite the proposed relation between sEBR and OSPAN as a biomarker of striatal dopamine, we found no relationship between these measures in either the cabergoline ($\tau_{(25)}$ = .06, p = .67) or placebo session ($\tau_{(25)}$ = .06, p = .67). This relationship was absent even when these measures were collected in the same session; namely, during screening ($\tau_{(25)}$ = -.11, p = .47; see Figure 6b). Finally, alertness, as indicated by our simple RT task also did not differ between the placebo (M = 224 ms, SD = 15.8) and cabergoline condition (M = 225 ms, SD = 15.1) ($t_{(25)}$ = -.54, p = .59).

Thus, cabergoline did not affect sEBR or our objective measure of alertness.

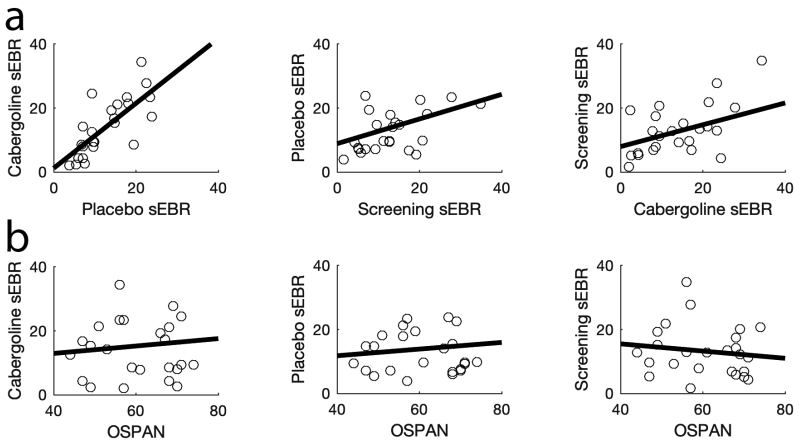


Figure 6. The relationship between sEBR measures among themselves and in relation to OSPAN. All sEBR measures were found to be positively correlated (**a**), but we found no relationship between sEBR and OSPAN (**b**)

Main experimental paradigms

We next examined potential effects of cabergoline on our main experimental measures of interest: target identification accuracy and processing in the backward masking task, attentional blink size in the attentional blink task, and

discrimination accuracy under varying conditions of difficulty and probability in the probabilistic discrimination task. To foreshadow our results, cabergoline did not affect any of the key behavioral findings associated with these tasks, whether sEBR or OSPAN were included as covariates in the analyses or not. Neither did it affect neural processing of the target in the backward masking task, as shown by ERP analyses. Yet, importantly, we did replicate all standard findings typically obtained with these tasks (e.g., effects of masking on target-evoked ERPs, the attentional blink).

Backward masking

Behavior. As is typically observed (e.g., Breitmeyer, 2007) and shown in Figure 7, targets were more often identified correctly in the backward masking task as the delay between target and mask (SOA) increased ($F_{(4, 100)} = 187.4$, $p < .001$, $\eta^2 = .79$, $BF_{incl} > 100$). Yet, in contrast to our main prediction that cabergoline would affect participant's ability to detect targets, there was no interaction between Drug and SOA ($F_{(4, 100)} < 1$, $p = .56$; $BF_{incl} = .03$; see Figure 7). We also did not find an overall difference between placebo and cabergoline on target identification accuracy ($F_{(1, 25)} < 1$, $p = .85$, $BF_{incl} = .13$). Controlling for screening sEBR, OSPAN, or the combination of both did not change the results. Neither did the inclusion of a between-subject factor for drug order. We repeated these analyses with d' as the dependent variable, but results were equivalent.

Similarly, in the case of confidence scores, cabergoline did not affect reported confidence ratings, as indicated by the lack of an interaction between SOA and Drug ($F_{(4, 100)} = 1.6$, $p = .21$; $BF_{incl} = .06$; see Figure 7). We also found no overall difference between drug conditions ($F_{(1, 25)} < 1$, $p = .63$; $BF_{incl} = .14$), but participants did report an improvement in response confidence as SOA increased ($F_{(4, 100)} = 204.8$, $p < .001$, $\eta^2 = .8$; $BF_{incl} > 100$). Again, correcting for baseline dopamine measures or drug order had no impact on the results. As with d', repeating these analyses with meta-d' as the dependent variable made no difference. Thus, we replicated a similar pattern of results with regards to both objective and subjective aspects of the participant's response during backward masking (Del Cul et al., 2007), but these patterns were not affected by cabergoline.

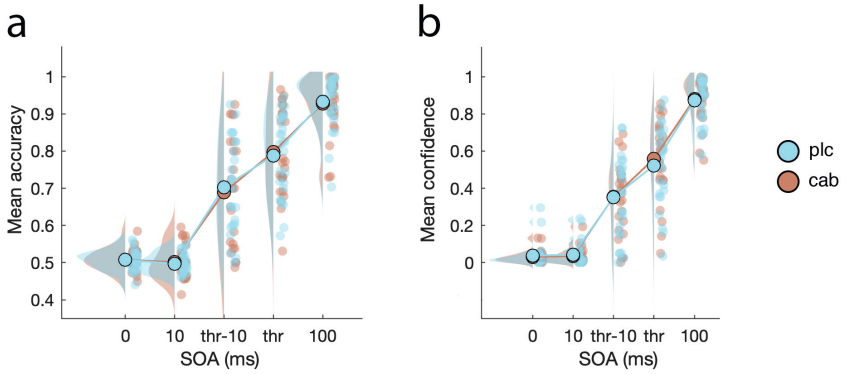


Figure 7. Mean accuracy and confidence scores in the backward masking task across five possible SOAs and both drug conditions. While both accuracy (a) and confidence (b) scaled with the duration of SOA, cabergoline had no effect on either of these measures

EEG. In addition to these behavioral masking effects, we replicated previous reports that target-evoked ERP components scale with the duration of target-mask SOA (Del Cul et al., 2007). We found a strong effect of SOA on the amplitude and peak-latency of the target-evoked P1, N1, N2, and P3b components. However, in the case of our ERP results as well, each of these measures was unaffected by cabergoline (see Table 2 and Figure 8). In order to remain close to previous studies (Del Cul et al., 2007), we repeated these analyses with an average reference instead of an earlobe reference, but the conclusions belonging to these results remained the same. Thus, cabergoline did not affect the threshold for conscious perception by modulating neural target processing.

Table 2. *Summary of the statistical analyses performed on ERP data.*

Component	Measure	SOA ($F_{(3, 75)}$ =)	Drug ($F_{(1, 25)}$ =)	SOA*Drug ($F_{(3, 75)}$ =)
P1	Amplitude	51.7, p < .001*	.01, p = .91, BF_{incl} = .16	.62, p = .60, BF_{incl} = .08
	Latency	8.2, p < .001*	.17, p = .68, BF_{incl} = .16	.48, p = .7, BF_{incl} = .08
N1	Amplitude	12.4, p < .001*	.04, p = .85, BF_{incl} = .15	.44, p = .73, BF_{incl} = .06
	Latency	66.8, p < .001*	.33, p = .57, BF_{incl} = .19	.07, p = .98, BF_{incl} = .06
N2	Amplitude	4.7, p = .005*	.04, p = .83, BF_{incl} = .15	.21, p = .89, BF_{incl} = .06
	Latency	99.3, p < .001*	2.9, p = .1, BF_{incl} = .38	.09, p = .97, BF_{incl} = .06
P3b	Amplitude	57.2, p < .001*	.02, p = .9, BF_{incl} = .15	.9, p = .44, BF_{incl} = .07
	Latency	10, p < .001*	.1, p = .74, BF_{incl} = .15	1.7, p = .18, BF_{incl} = .09

* = BF_{incl} > 100

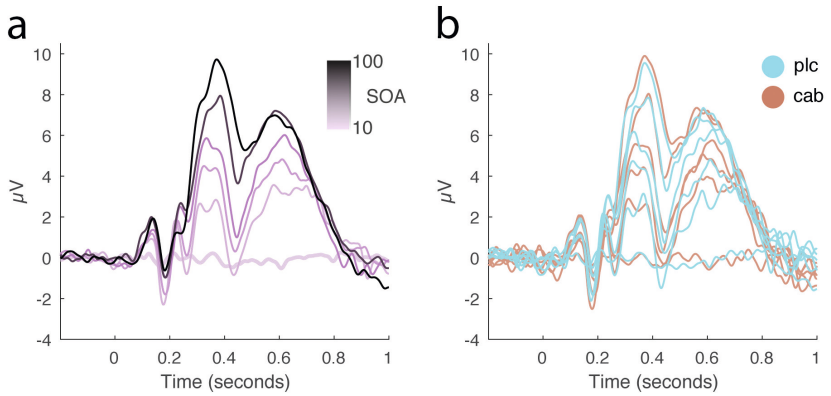


Figure 8. Effects of cabergoline on target-evoked ERPs in the backward masking task. This figure displays the grand-average target-evoked ERPs for P3b electrodes (P1, Pz, P2, PO3, POz, PO4) for both drug conditions combined **(a)** and separately **(b)**, per target-mask SOA condition (10, 40, 50, 60, 70, and 100 ms). This figure shows that while ERP amplitudes and latencies generally increased as a function of target-mask SOA, these measures were unaffected by cabergoline.

Attentional blink

As expected and shown in Figure 9, we found a robust attentional blink: T2|T1 accuracy was significantly worse when T2 was presented after one distractor (lag 2) compared to seven distractors (lag 8; $F_{(1, 24)} = 43.9$, $p < .001$, $\eta^2 = .39$; $BF_{incl} > 100$). T2|T1 accuracy was marginally better in the cabergoline condition (81.9±11%) compared to placebo (79.7±14%; $F_{(1, 24)} = 4.2$, $p = .052$; $BF_{incl} = .34$), but we found no interaction between Lag and Drug ($F_{(1, 24)} = 1.42$, $p = .25$; $BF_{incl} = .3$).

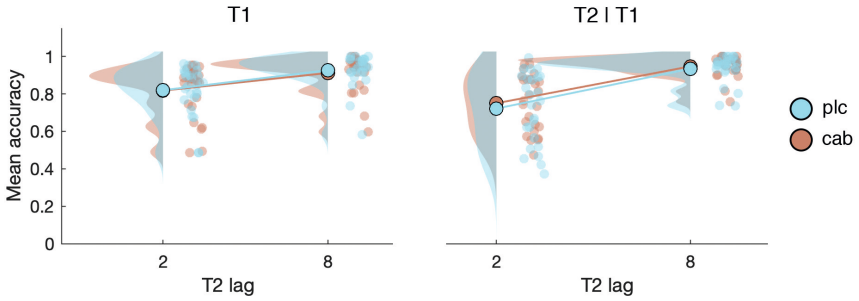


Figure 9. Mean accuracy for T1 and T2|T1 at lag 2 and lag 8. While we replicated behavioral findings common to the attentional blink paradigm, cabergoline did not affect the attentional blink

When we controlled for baseline dopamine measures (i.e., baseline OSPAN and sEBR), the inclusion of OSPAN did not affect the results, and controlling for sEBR also did not reveal a critical interaction between Lag and Drug ($F_{(1, 22)} = 1.3$, $p = .26$; $BF_{incl} = .3$). We did find a between-subject effect of screening

sEBR ($F_{(1, 22)}$ = 5.2, p = .03; BF_{incl} = 2.5), which exacerbated the main effect of Drug ($F_{(1, 22)}$ = 5.4, p = .029; η^2 = .004; BF_{incl} = .36). In addition, we found an interaction between sEBR and Drug ($F_{(1, 22)}$ = 5.8, p = .025; η^2 = .005): sEBR correlated with overall T2|T1 accuracy in the placebo ($\tau_{(25)}$ = -.31, p = .03), but slightly less so in the cabergoline ($\tau_{(25)}$ = -.26, p = .07) condition.

Although the critical three-way interaction between sEBR, Lag and Drug was not significant, we also found an interaction between screening sEBR and Lag ($F_{(1, 22)}$ = 5, p = .036; η^2 = .004). This interaction is best understood in terms of the relationship between sEBR and AB size (Colzato et al., 2008; Slagter et al., 2012); namely, the difference in T2|T1 accuracy between lag 2 and 8. AB size ranged from -2.9% to 59.7% in the placebo condition, and from 0% to 47.5% in the cabergoline condition. The interaction between sEBR and Lag stems from a small correlation between screening sEBR and AB size in the cabergoline ($\tau_{(25)}$ = .31, p = .03) and placebo condition ($\tau_{(25)}$ = .28, p = .047). We found no other relationships between dopamine baseline measures and the AB; the difference in AB size between the cabergoline and placebo session was unrelated to screening sEBR ($\tau_{(25)}$ = .05, p = .71). For the cabergoline session, there was no relation between AB size and sEBR ($\tau_{(25)}$ = .13, p = .35), nor between AB size and OSPAN ($\tau_{(25)}$ = .08, p = .57; see Figure 10a). Similarly, for the placebo session we did not find a relationship either between AB size and sEBR ($\tau_{(25)}$ = .15, p = .3), nor OSPAN ($\tau_{(25)}$ = .02, p = .89, see Figure 10b). However, none of the correlations reported in this section survived a Bonferroni correction for multiple comparisons using a corrected alpha level of .05 / 9 = .0056, based on all correlations computed in this section. Thus, we found no strong support for a relationship between sEBR and the AB, or OSPAN and the AB, nor did controlling for these factors reveal drug-related effects on the AB.

When controlling for drug order, we found an interaction between Drug and drug Order ($F_{(1, 21)}$ = 5.7, p = .027; η^2 = .004; BF_{incl} = .43). When participants received placebo first, they were less accurate overall in identifying T2|T1 in the placebo condition (77.5±15%), compared to the cabergoline condition (81.6±11%). For participants who received cabergoline first, this difference was not present (placebo: 82.1±12%, cabergoline: 82.3±10%). No other order effects were present. Together these results indicate that cabergoline did not affect the attentional blink.

We found no difference in T1 accuracy between the placebo and cabergoline condition ($F_{(1, 24)}$ < 1, p = .49; BF_{incl} = .27), or an interaction between Drug and Lag ($F_{(1, 24)}$ < 1, p = .5; BF_{incl} = .32), but T1 accuracy was worse when T2 was presented after only one distractor (lag 2; $F_{(1, 24)}$ = 46.6, p < .001, η^2 = .16; BF_{incl} >

100). Controlling for sEBR, OSPAN, or drug order did not affect the results for T1 accuracy. Thus, cabergoline also did not affect T1 identification.

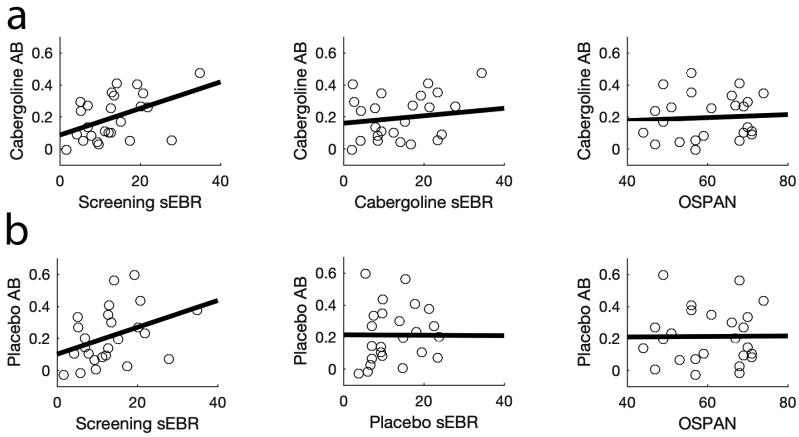


Figure 10. The relationship between AB size and striatal-dopamine proxy measures. Both for the cabergoline and placebo condition, we found a positive relation between AB size and screening sEBR, while such a relationship was not present neither for sEBR measured within the same session, nor for OSPAN

Probabilistic discrimination

In our final experimental paradigm, participants were tasked with discriminating between face and house stimuli that varied in difficulty and probability of occurrence (see Methods). Cabergoline did not affect discrimination accuracy (specifically: hit-rate; $F_{(1, 23)} = 2.4$, $p = .13$, $BF_{incl} = .35$). Neither did we find an interaction between Drug and Difficulty ($F_{(1, 23)} < 1$, $p = .91$, $BF_{incl} = .14$), Drug and Probability ($F_{(4, 92)} < 1$, $p = .78$, $BF_{incl} = .02$), or a three-way interaction ($F_{(4, 92)} < 1$, $p = .48$, $BF_{incl} = .05$; see Figure 11a). Thus, cabergoline did not significantly affect perception on this task either. Standard effects observed with this task in non-drug studies were replicated (Bauer et al., 2016, September): accuracy was lower on difficult ($F_{(1, 23)} = 208.3$, $p < .001$, $\eta^2 = .60$, $BF_{incl} > 100$), and on low-probability trials ($F_{(4, 92)} = 22.49$, $p < .001$, $\eta^2 = .04$, $BF_{incl} > 100$). When we controlled for screening sEBR and OSPAN, this did not change the results. We did find an interaction between drug Order and Drug ($F_{(4, 88)} = 5.2$, $p = .03$, $BF_{incl} = 1.3$). Opposite to what we found for the attentional blink task, when participants received placebo first, they were more accurate overall in the placebo (84±7%) compared to the cabergoline (82.4±7%) condition. When cabergoline was administered first, this difference was absent (placebo: 83.5±5%, cabergoline: 83.1±4%). We found no other order effects.

When we repeated these analyses for participants with all three difficulty levels for both stimuli after titration (N = 16; see Methods), the same pattern of results was obtained. Cabergoline did not affect discrimination accuracy ($F_{(1, 15)} < 1$, $p = .47$, $BF_{incl} = .69$), and we found no two-way interaction between Drug and Difficulty ($F_{(2, 30)} < 1$, $p = .59$, $BF_{incl} = .05$), Drug and Probability ($F_{(4, 60)} < 1$, $p = .76$, $BF_{incl} = .02$), or a three-way interaction ($F_{(8, 120)} = 1.1$, $p = .4$, $BF_{incl} = .04$; see Figure 11b). Accuracy was still lower on difficult ($F_{(2, 30)} = 233.8$, $p < .001$, $\eta^2 = .53$, $BF_{incl} > 100$) and low-probability trials ($F_{(4, 60)} = 21.1$, $p < .001$, $\eta^2 = .08$, $BF_{incl} > 100$). Controlling for screening sEBR and OSPAN did not change the acquired results. The interaction between drug Order and Drug was maintained ($F_{(1, 14)} = 9.9$, $p = .007$, $BF_{incl} > 100$), where overall accuracy was higher in the placebo condition (83±7%) when placebo was administered first compared to the cabergoline condition (80.6±6%). This difference was again absent when cabergoline was administered first (placebo: 82.7±3%, cabergoline: 83.1±3%).

Thus, we found no evidence in support of the conclusion that cabergoline affected any of the key findings in our perceptual tasks.

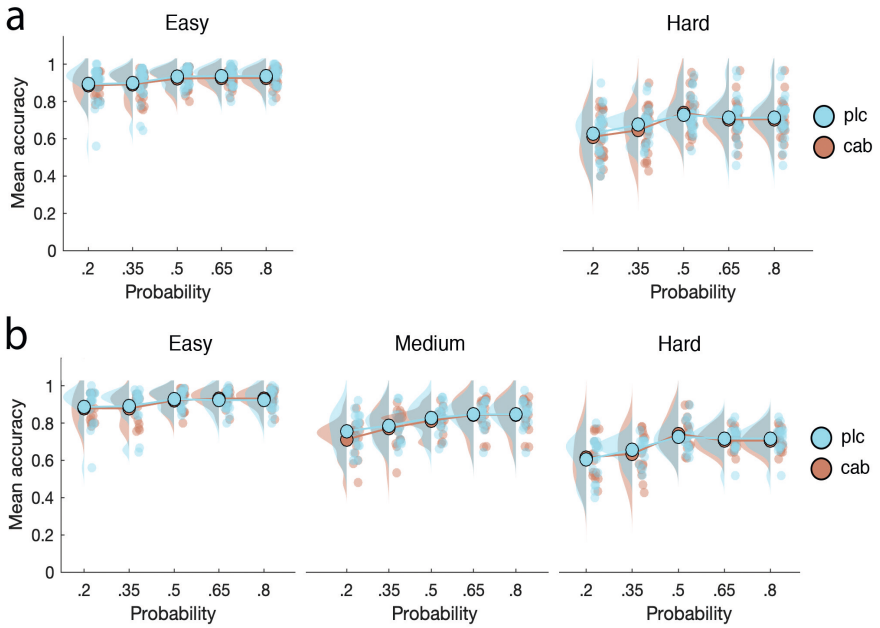


Figure 11. Cabergoline did not affect probabilistic discrimination performance. This figure shows mean accuracy for each drug condition and across five stimulus probabilities for all participants with two (**a**; N = 24) and three (**b**; N = 16) difficulty levels after titration. While accuracy was higher for easy trials and blocks in which the presented stimulus was more likely to occur, these effects were unaffected by cabergoline. Medium-difficulty trials take up position in between easy and hard trials in terms of mean accuracy when included

2.5. Discussion

This double blind, placebo-controlled, cross-over study tested the hypothesis that striatal dopamine is involved in conscious perception by administering the dopamine D2 agonist cabergoline and placebo to healthy participants. To that end 1) we established an effect of cabergoline on participant's physiological and subjective state, 2) we tested the effect of cabergoline on two often-used dopamine proxy measures (sEBR and OSPAN), and 3) we subjected participants to well-known and often-employed experimental paradigms targeting the neural correlates of consciousness. While we were able to establish an effect of cabergoline on participant's physiological and subjective state, we did not find an effect of cabergoline on sEBR. Crucially, while we replicated key behavioral and ERP findings associated with the paradigms we employed (Del Cul et al., 2007; Slagter et al., 2017; Van Opstal et al., 2014), none of these findings were affected by cabergoline. Thus, we did not obtain evidence that the dopamine D2 agonist cabergoline affected conscious perception, which was also supported by Bayesian statistics.

Just as with positive results, an observed null result could indicate a true effect (i.e., no role for striatal dopamine in conscious perception), or it could be due to uncontrolled or unknown factors. Based on the convergent evidence discussed in the introduction implicating the striatum and its irrigation by dopamine in conscious perception (e.g., Bisenius et al., 2015; Slagter et al., 2017; Van Opstal et al., 2014), we predicted that cabergoline would affect performance on the backward masking and attentional blink tasks. Yet, our Bayesian results provided evidence against an effect of 1.5mg cabergoline (all $BF_{incl} < .7$; except for the interaction between Drug and drug Order for discrimination accuracy in the probabilistic discrimination task). This absence of an effect is further strengthened by the fact that other studies using different cognitive tasks have previously reported significant effects of cabergoline on task performance using a similar or an even lower dose of orally administered cabergoline (Broadway et al., 2018; Cavanagh et al., 2014; Fallon et al., 2017; Frank & O'Reilly, 2006; Norbury et al., 2013; Nandam et al., 2013; Yousif et al., 2016). These observations together argue for an interpretation in terms of a true null result. Nonetheless, there are a number of practical limitations that pertain to the present study.

For one, blinding was unsuccessful as 85% of participants guessed when they received cabergoline at the end of the experiment. While all participants with adverse reactions were excluded from any analyses, we cannot exclude the possibility that participants' ability to tell when they received cabergoline influenced the results. It is possible this ability stems from the higher dosage

(1.5 mg) we employed compared to previous studies administering cabergoline (1-1.25 mg; Broadway et al., 2018; Cavanagh et al., 2014; Frank & O'Reilly, 2006; Norbury et al., 2013; Nandam et al., 2013; Yousif et al., 2016). While a lower dosage may benefit blinding, it may also further reduce the chance of finding an effect of cabergoline. Another option for future research would be to administer an antiemetic such as domperidone in combination with cabergoline, in order to mitigate and mask physical side effects (Fallon et al., 2017; Norbury et al., 2013).

Second, while we titrated performance on the backward masking and probabilistic discrimination tasks, we did not do so for the attentional blink task, which may have resulted in reduced sensitivity of this task to our dopamine manipulation. We chose this specific version of the task in order to stay as close as possible to a recent study in which we found a relationship between activity in the ventral striatum and the attentional blink (Slagter et al., 2017). While we cannot rule out that a titration procedure may have resulted in a task more sensitive to our manipulation, we believe this is unlikely given that our participants showed a robust attentional blink, indicating that our task was suitable for the effect we intended to manipulate. In addition, since cabergoline did not affect performance on the two tasks we did titrate, we believe it unlikely that this would have been different in the case of the attentional blink task.

Third, a limitation of our study pertains to the use of alleged indirect measures of dopamine activity: sEBR (Jongkees & Colzato, 2016) and the OSPAN task (Cools, Gibbs, Miyakawa, Jagust, & D'Esposito, 2008). The functioning of dopamine relies on a relative equilibrium at the system level (i.e. frontal versus basal ganglia; Wiecki & Frank, 2013; Wise, Murray, & Gerfen, 1996), within the basal ganglia itself (i.e., direct versus indirect pathways; Redgrave et al., 2010), and at the level of synapses (i.e., pre- versus postsynaptic effects; Frank & O'Reilly, 2006; Usiello et al., 2000). The equilibrium at the system- and pathway-level is in turn implicated in the differing effects at the receptor level (i.e., D1 versus D2 receptors; Shen, Flajolet, Greengard, & Surmeier, 2008). Due to the complexity and the multitude of interactions between different systems it is especially important to be able to report on the initial state of the dopamine-system.

Unfortunately, sEBR was unaffected by cabergoline. While this finding goes against studies with humans showing effects of dopaminergic drugs on sEBR (Blin, Masson, Azulay, Fondarai, & Serratrice, 1990; Karson, 1983; Strakowski & Sax, 1998), this finding fits with several pharmacological studies that reported no effect (Depue, Luciana, Arbisi, Collins, & Leon, 1994; Ebert, 1996; Mohr, Sándor, Landis, Fathi, & Brugger, 2005; van der Post, de Waal, de Kam, Cohen, & van Gerven, 2004).

Moreover, differences in OSPAN and sEBR were unpredictive for the effect of cabergoline on performance in our experimental paradigms. These findings are surprising in light of previous work showing how effects of cabergoline on cognitive function depend on individual baseline sEBR (Cavanagh et al., 2014) and OSPAN score (Broadway et al., 2018). However, it should be emphasized that evidence for a relationship between dopamine-proxy measures and dopamine levels is often correlational, based on studies with small sample sizes ($N < 50$; Cools et al., 2008), and that results are mixed (Dang et al., 2017; Sescousse et al., 2018). It is debatable whether such small-sample studies have adequate power to provide evidence for this relationship (Cremers, Wager, & Yarkoni, 2017; Rousselet & Pernet, 2012). Similarly, the correlational findings we report should also be interpreted cautiously. While measures such as OSPAN and sEBR are easy to administer, the correlational evidence – as well as inconsistent findings regarding these measures (Dang et al., 2017; Jongkees & Colzato, 2016; Sescousse et al., 2018) – put pressure on their validity as an index of baseline dopamine levels.

Despite these practical limitations, the advantage of our study is that our results are univocal: the administration of cabergoline had no effect on any of our indices of conscious perception obtained using often-employed experimental paradigms in the study of consciousness. Of importance, we did consistently replicate all key behavioral and event-related potential (ERP) findings associated with each of these tasks.

Theoretically, perhaps our null-findings are best explained by an appeal to predictive processing approaches to brain function (Hohwy, 2012; 2013; de Lange, Heilbron, & Kok, 2018). From the perspective of these theories, the brain is viewed as a prediction machine which is continuously revising its predictions about the causes of sensory data. On this view, top-down signals in perceptual hierarchies carry perceptual predictions, while bottom-up signals convey perceptual prediction errors. Perception becomes a process of continual minimization of prediction errors across hierarchical levels, instantiating a process of approximate Bayesian inference on the causes of sensory signals. In this process, dopamine is thought to regulate the relative precision of top-down predictions and bottom-up prediction errors, where precision is understood as inverse variance (Friston et al., 2012): a higher (expected) variance of sensory signals leads to a smaller influence on updating perceptual predictions. "If true, this means that modulators of synaptic gain (like dopamine) do not report perceptual content but the context in which percepts are formed. In other words, dopamine reports the precision or salience of sensorimotor constructs

(representations) encoded by the activity of the synapses they modulate." (Friston et al., 2012, p. 2).

From this perspective, there may not have been enough uncertainty in our experimental paradigms for our dopamine manipulation to play a determining role. Indeed, previous studies employing cabergoline in the context of working memory found an effect under conditions where task-relevant targets were embedded in a larger perceptual field; namely, in the presence of distractors. In these studies, cabergoline exerted an effect on task performance in terms of target-detection accuracy and successful recall when a target stood in competition with other stimuli (Broadway et al., 2018; Cavanagh et al., 2014; Frank & O'Reilly, 2006). In the case of each of our tasks, targets were consistently surrounded by distracting information *temporally, but never spatially.* Perhaps the selective functionalities of the basal ganglia and dopaminergic firing actualize primarily when the system is under pressure to select sensorimotor constructs in the face of multidimensional uncertainty.

Our data cast doubt on a causal role for dopamine in visual perception, as shown by a significant lack of effect in several standard perceptual paradigms. Future studies, perhaps with more direct measures of dopaminergic activity, and more naturalistic paradigms including more opportunities for selection and/or uncertainty, may yet reveal a more specific influence of this neurotransmitter on how we perceptually encounter the world.

Funding information
This work was supported by an ERC starting grant to HAS by the H2020 European Research Council [ERC-2015-STG-679399] and an Amsterdam Brain and Cognition (ABC) grant to HAS, AKS and CSL. AKS is additionally grateful to the Canadian Institute for Advanced Research (CIFAR) Azrieli Programme on Brain, Mind, and Consciousness, and to the Dr. Mortimer and Theresa Sackler Foundation.

Chapter 3

Conscious perception and the role of the basal ganglia: preliminary findings from a deep brain stimulation study

---

In preparation as:

Boonstra E. A., Bais, M. N., van Schouwenburg M. R., van den Munckhof, P., Smit, D. J. A., Denys, D., & Slagter, H.A. (in prep.). Conscious perception and the role of the basal ganglia: preliminary findings from a deep brain stimulation study.

3.1. Abstract

Conscious perception is thought to depend on global amplification of sensory input. In recent years, the basal ganglia have been implicated in gating conscious access due to their consistent involvement in thalamocortical loops. However, much of the evidence implicating the basal ganglia in these processes in humans is correlational. The current study is a preliminary investigation in four patients to explore whether deep brain stimulation (DBS) in the basal ganglia might improve conscious perception. In our study, treatment-resistant obsessive-compulsive disorder (OCD) patients with a striatal DBS implant completed two canonical conscious perception tasks: emotion-induced blindness and backward masking. We found preliminary evidence in support of a role played by the basal ganglia in conscious perception at the behavioral level: patients performed better when stimulation was active, but we could not establish neural effects corresponding to these behavioral findings, possibly due to our small sample size. We discuss the potential implications and limitations of our study and delineate avenues for future research.

3.2. Introduction

The relationship between consciousness and the brain is often considered as one of the major frontiers of contemporary science. Within the context of conscious visual perception, one major outstanding question concerns how a stimulus becomes accessible for report. Influential theoretical accounts have proposed that such accessibility requires "broadcasting" of a stimulus to the whole brain through thalamocortical circuits (Crick & Koch, 2003; Dehaene & Changeux, 2011; Edelman, 2003). The basal ganglia (BG) are a cluster of subcortical nuclei located deep in the brain, which modulate these circuits (Smith et al., 2004), and as such, may gate conscious perception. For decades we have known that BG structures contribute to sensory and perceptual processes (Afrasiabi et al., 2021; Alexander & Crutcher, 1990; Arsalidou et al., 2013; Brown et al., 1997; Seger, 2013), in line with the notion that the BG not only perform action selection, but fulfill a general selection function, including selection for conscious perception (e.g., Redgrave et al., 1999). It is notable in this respect that fMRI studies consistently show differences in BOLD activity in the striatum between consciously perceived and unconscious stimuli using backward masking tasks (Bisenius et al. 2015). Another repeated finding is a relationship between striatal D2 receptor binding and performance on conscious perception tasks (Bisenius et al., 2015; Slagter et al., 2010, 2012; Van Opstal et al., 2014; but see Boonstra et al., 2020). Most of these studies employ neuroimaging, and thereby lack temporal precision and/or a direct measure of BG activity. This has left it largely unclear how the BG may contribute to conscious perception in humans.

In recent years, deep brain stimulation (DBS) electrodes have proven an effective tool to not only modulate BG structures, but also to measure their activity. DBS involves an invasive procedure leveraged to alleviate symptoms of several treatment-resistant neurological and psychiatric pathologies such as Parkinson's disease (Almeida et al., 2017), anorexia nervosa (Whiting et al., 2018), depression (Dandekar et al., 2018), and obsessive-compulsive disorder (OCD, Baldermann et al., 2021). A previous study employed an attentional blink task to probe conscious perception in patients with DBS electrodes implanted in the ventral striatum (Slagter et al., 2017). The attentional blink comprises a deficit in consciously perceiving the second of two targets (T1 and T2) whenever it follows the first target within 100–500 ms in a rapid stream of distractors (Raymond et al., 1992). Patients performed this task while intracranial EEG (iEEG) activity from the ventral striatum was recorded using the DBS electrodes without stimulation. Intracranial EEG recordings revealed a consistent burst in theta-band activity only when T2 was consciously perceived (Slagter et al.,

2017). Moreover, increased activity in α and low β frequencies after perceiving T1 was observed in trials in which subjects subsequently failed to perceive T2, possibly reflecting attentional capture by T1. These findings support the notion that the BG contribute to conscious experience. Yet, like neuroimaging, they only provide correlational evidence.

In the present study, we present preliminary data from a small sample of four treatment-resistant OCD patients with DBS electrodes implanted in the ventral striatum. The efficacy of DBS treatment in OCD patients is still under active investigation, but evidence thus far encourages further study (Acevedo et al., 2021; Graat et al., 2021; van Wingen et al., 2022; Wu et al., 2021). Our study goes beyond past work in that we were able to record iEEG during DBS. We were thereby able to causally manipulate and measure striatal activity simultaneously, while patients performed two canonical conscious perception paradigms: emotion-induced blindness (EIB) and backward masking (BM). This setup allowed us to establish a more conclusive role for the ventral striatum in conscious perception compared to the earlier study by our group. We concurrently measured scalp EEG, permitting investigation of the effects of DBS on common event-related potentials (ERPs) markers of conscious perception (see Hypotheses below). Moreover, the two tasks allowed us to probe a participant's stimulus detection ability both when the perception of a target is complicated by an emotionally valenced (positive or negative; EIB), as well as a neutral disturbance (BM). Taken together, this setup allowed us to examine if causal perturbation of striatal activity would affect conscious perception within the context of these tasks.

Hypotheses

Emotion-induced blindness refers to an impaired awareness of stimuli appearing soon after an irrelevant, emotionally arousing stimulus (Most et al., 2005). It is therefore an effect similar to the attentional blink, except that instead of T1, emotionally arousing or valence neutral distractors are presented 200-800 ms before the target. This manipulation of emotional arousal allowed us to vary the degree to which distractors may capture attention in an attempt to extend previous findings (Slagter et al., 2017).

Behaviorally, OFF DBS, we expected to find an effect akin to a standard attentional blink, based on past EIB studies (Kennedy et al., 2014; Most et al., 2005), as reflected in lower accuracy when the lag between the distractor and target is short rather than long. We expected this difference to be larger for targets preceded by negative distractors compared to neutral distractors due to increased attentional capture. In terms of ERPs derived from scalp EEG data, a

usual finding for the EIB task is that N2 and P3 components are suppressed for targets following an emotionally arousing distractor (Kennedy et al., 2014). We expected to replicate these findings for targets following a negative distractor OFF DBS.

In the backward masking task, the order of distractor and target is reversed. A target is presented first, followed by a mask (akin to a distractor) at a variable delay, known as the stimulus onset asynchrony (SOA). In this task, patients indicated whether the target, a number, was smaller or larger than five. Commonly, accuracy is better when the delay between target and mask is long rather than short (Breitmeyer, 2007). Behaviorally, OFF DBS, we expected to replicate this finding in terms of improved accuracy for longer SOAs. Our task departs from the standard version in that patients do not only report on the target, but simultaneously indicate the confidence in their response (sure/unsure). Like accuracy, we expected response confidence to improve with increased SOA duration. The reason to include a confidence measure is that the ventral striatum is thought to be involved in determining self-confidence (Kiverstein et al., 2019). We expected DBS ON to improve patients' accuracy and response confidence. In terms of scalp ERPs, increased ERP amplitude is usually found as SOA increases (Del Cul et al., 2007).

Our study is novel in that we could investigate changes in neural oscillatory activity recorded from the ventral striatum within the context of an EIB and BM task ON and OFF DBS. Based on our previous iEEG attentional blink study (Slagter et al., 2017), OFF DBS, we expected that only consciously perceived targets would be associated with an increase in θ activity between 150 and 400 ms after stimulus onset. In the EIB task, we also expected that failure to perceive the target would be foreshadowed by a distractor-induced increase in α and low β oscillatory activity. Moreover, in the case of EIB, we expected DBS ON to reduce capture by negative distractors and thereby to improve accuracy. At the neural level, we expected this change to be reflected in a reduced activity in α and early β frequencies immediately following the distractor, as well as an increased θ burst and larger N2 and P3 amplitudes to seen targets. In the BM task, we similarly expected the stimulation to increase induced θ activity and N2 and P3 amplitudes to seen targets.

Below, we present preliminary findings against the background of earlier work suggesting that DBS may alter patient's capacity for conscious perception. However, given our small sample size of only four patients, we present solely descriptive statistics and graphical analysis, necessitating more research to substantiate these findings.

3.3. Methods

Participants

Eleven therapy-resistant OCD patients eligible for deep brain stimulation (DBS) were recruited from the outpatient clinic for DBS at the department of Psychiatry of the Amsterdam University Medical Center, location Academic Medical Center (AMC) from 2015-2020. Alcohol or substance abuse in the past 6 months was a reason for exclusion from the study. Out of the eleven patients recruited, four participated in the tasks comprising the present study (mean age: 48 years, range: 35-61, 3 women, 1 man). Not all eleven patients completed the tasks of the current study because initially the study protocol spanned a full year with the current tasks at the end, for which it turned out the battery life of the stimulator was not suited. One patient performed at chance level in the backward masking (BM) task and was thus excluded from further analyses, resulting in a sample of three patients for this task (mean age: 43 years, range: 35-54, 3 women). All participants were right-handed and took their standard medications. Medications included clomipramine (37.5-75 mg), quetiapine (300 mg), oxazepam (10 mg), fluvoxamine (25-100 mg), fluoxetine (60 mg), lithium carbonate (1000 mg), mirtazapine (45 mg), olanzapine (15 mg), promethazine (25 mg), and lorazepam (1.5 mg). The study was approved by the medical ethics board of the AMC and all patients provided written consent to participate. The trial was registered in the Netherlands Trial Register (Trial NL7486). Patients also performed several other tasks on different study visits (data to be published elsewhere, e.g., Fridgeirsson et al., 2021).

Experimental Paradigms

Stimuli were presented at a distance of 70 cm on a 15.4-inch laptop screen (HP 6730b, refresh rate = 60 Hz, resolution 1024x768). Due to limited storage capacity of the DBS recording device, and to maximize the number of trials, patients completed a variable number of trials on each block for both tasks.

Emotion-Induced Blindness

The emotion-induced blindness task consisted of a central rapid serial visual presentation of briefly (100 ms) presented pictures of landscapes (6.5° wide and 4.9° high) against a black background. Participants were instructed to identify target in this stream (a rotated landscape 90°), and to indicate whether this target was rotated left or right (Fig. 1). As a third response option they could indicate no rotated picture was present (no target). Before the presentation of a target, which was present in each trial, on approximately 86% of trials, a task-

irrelevant distractor image was shown, which could be either a neutral image or a negative image, or on approximately 14% of trials, there was no distractor at all. The valence and arousal of these images were rated by the participants themselves on a scale from 1 to 9 on a separate occasion. The 40 most negatively rated pictures served as negative distractors. As neutral distractors we selected 40 pictures rated both least arousing and least positive (but not negative). The target followed the distractor either after 200 ms or 800 ms. Each trial started with an intertrial interval of 200 ms, after which a stimulus stream consisting of 17 stimuli began. A white central fixation cross was visible throughout the experiment. All stimuli were presented using Presentation (version 14.5; Neurobehavioural Systems) at the center of the screen.

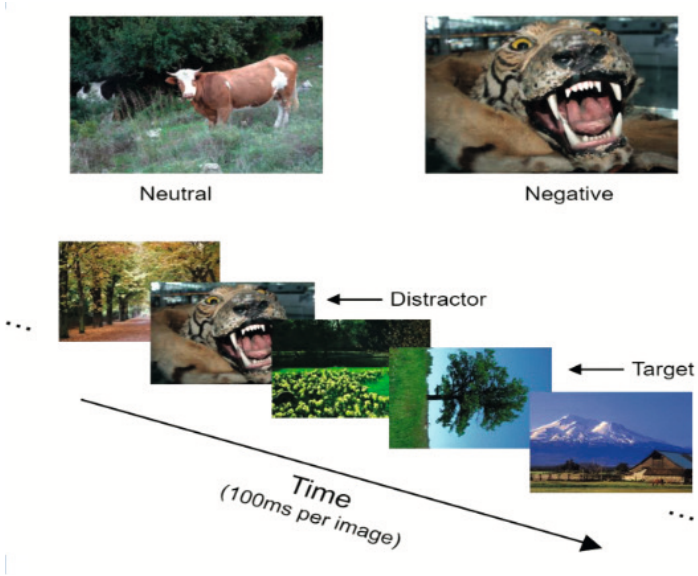


Figure 1. Experimental procedure for one trial of the emotion-induced blindness task

Backward Masking

The backward masking task was adapted from Van Opstal and colleagues (2014; see Fig. 2). Participants indicated whether briefly presented masked digits (1, 4, 6, or 9) were smaller or larger than 5 and simultaneously rated their confidence in their response, resulting in four response options (>5 sure, >5 unsure, <5 sure, <5 unsure). Responses were counterbalanced in terms of left and right across participants. Each trial started with the presentation of a central fixation cross (53 point Courier New), which increased in size (106 point Courier New, 150 ms duration), cueing the impending target. The target stimulus (53 point Courier

New) then appeared for 16 ms at one of two positions centered at the vertical midline (top or bottom, 3.9° from fixation). Both stimulus locations were equally probable. A mask followed the target (200 ms duration) at a variable stimulus onset asynchrony (SOA), with a duration of 16, 33, 50, 66, or 100 ms. By varying the delay between cue and target dependent on SOA, the delay between cue and mask was held constant at 800 ms. The mask (53 point Courier New) was composed of two letters "E" and two letters "M", tightly surrounding the target location without superimposing or touching it. All stimuli were black and presented on a white background, using the Psychophysics toolbox for MATLAB (Brainard, 1997). The central fixation cross was visible throughout the experiment.

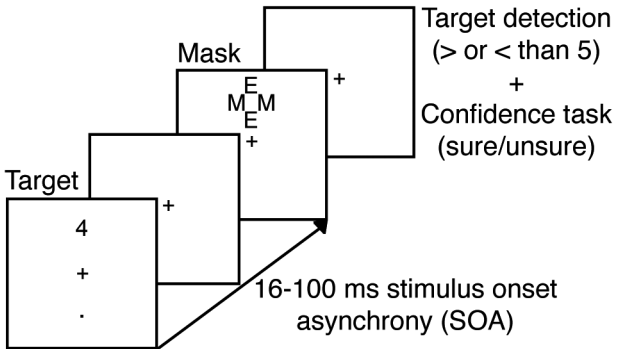


Figure 2. Experimental procedure for one trial of the backward masking task (adapted from Van Opstal et al. 2014)

Neural data recording and preprocessing

Preprocessing was done using the Fieldtrip toolbox (Oostenveld et al., 2011) for Matlab (The MathWorks, Inc. Natick, MA, USA) using custom-written scripts.

DBS and Intracranial EEG

Two quadripolar electrodes (Model 3389 Medtronics) with 4 contact points of 1.5 mm each separated by 0.5 mm were implanted stereotactically. Implantation was performed bilaterally through the anterior limb of the internal capsule with the deepest, i.e. most ventral, contact point located in the nucleus accumbens, 3 mm anterior to the anterior commissure and the 3 upper contact points positioned in the ventral part of the anterior limb of the internal capsule (vALIC). The neurostimulator used was the Activa PC + S (Medtronics; model 37604) which allows both for stimulating and recording local field potentials (LFPs) with the recording and stimulation part of the device being completely

separated. Stimulation was performed through the middle two contact points at 130 Hz.

Intracranial EEG was recorded from the deepest contact point in the nucleus accumbens using the 8180 Sensing Programmer SW (Medtronic Inc.) with a sampling rate of 422 Hz Using hardware 0.5 Hz high pas filtering and 100 Hz low pass filtering. LFP data were high-pass "firws" filtered (default settings) at 0.5 Hz using a Kaiser window (Widmann et al., 2015). The continuous data were subsequently epoched from -1.5 to 2 s around stimulus presentation to facilitate time-frequency decomposition and baseline corrected to the average activity between -500 ms and -200 ms pre-stimulus in line with Slagter and colleagues (2017). Epochs containing eye blinks surrounding stimulus presentation were rejected based on the visual inspection of the scalp EEG data, that was concurrently measured.

Scalp EEG

EEG was recorded at a sampling rate of 1000 Hz using a 64-channel recording system with shielded Ag/AgCl electrodes (ANT Neuro B.V.) following the international 10–10 system. EEG data were high-pass "firws" filtered (default settings) at 0.1 Hz using a Kaiser window (Widmann et al. 2015) and low-pass filtered at 50 Hz due to noise from the DBS implant. Data were offline referenced to the average of all electrodes except LFP, EOG, and mastoid electrodes. The continuous data were subsequently epoched from -1.5 to 1.5 s around stimulus presentation and baseline corrected to the average activity between -500 ms and 200 ms pre-stimulus. We used this time window for both tasks to exclude the distractor in our baseline window in the EIB task, and to remain consistent with iEEG time-frequency analyses. Epochs containing EMG artifacts or eye blinks surrounding stimulus presentation were rejected based on visual inspection. Remaining eye blink artifacts were removed by decomposing the EEG data into independent sources of brain activity using independent component analysis, and removing eye blink components from the data for each subject individually. Epochs were low-pass filtered at 30 Hz for visualization purposes.

Data analysis

The small sample size precluded the use of inferential statistics; therefore, data were analyzed using descriptive statistics and graphic analysis.

Emotion-induced blindnes

Three out of four patients completed the task OFF DBS first. Our dependent measure is an accuracy measure reflecting the percentage of correctly identified

picture orientation out of all trials (including trials where patients reported no target). Due to our small sample size, relatively low signal-to-noise due to overlapping responses to the RSVP stimuli, and the absence of catch trials, we were unable to compute standard ERPs for this task and only report results of iEEG time-frequency analyses.

Backward masking

Two out of three patients completed the BM task OFF DBS first. Our two dependent measures were accuracy (correctly indicating whether the presented number was smaller or larger than 5) and confidence (whether patients were sure or unsure about their response).

To extract event-related potential (ERP) markers of information-processing, we epoched the ERP data to the onset of the mask Del Cul et al (2007). Next, we subtracted the data from the mask-only SOA condition from all other SOA conditions. Finally, we shifted ERP onset back to target onset, to compute target-locked ERPs. Due to the high levels of noise in the data, we focused solely on graphical analysis of the P3b over central parietal scalp sites (P1, Pz, P2, PO3, POz, PO4). This scalp site was used to determine the peak amplitude of this component by averaging the amplitude values 15 ms around the peak sample between 200-400 ms post-target.

3.4. Results

Emotion-induced blindness

Behavior

As shown in Fig. 3, as typically found (Shapiro et al., 1997), patients showcased a robust blink: they less often detected the target on Lag2 (M: 59.4%, SD: 6.6%) compared to Lag8 (M: 68.4%, SD: 10%) distractor-present trials. Moreover, the blink was larger after a negative versus a neutral distractor. That is, within Lag8 trials, patients performed better on average when presented with a negative (M: 70.8%, SD: 11.5%) compared to a neutral distractor (M: 65.9%, SD: 10.4%), while this seeming difference was smaller for Lag2 trials (NEG-M: 60.5%, SD: 7.5%; NEUT-M: 58.4%, SD: 7.9%), resulting in a numerically deeper blink (Lag8-Lag2) after a negative vs. neutral distractor: the emotion-induced blindness effect. Notably, in line with our prediction that DBS would reduce the distractor-induced blink, patients showcased improved accuracy when DBS was ON (M: 63.3%, SD: 6.5%) compared to when stimulation was OFF (M: 55.6%, SD: 4.5%) in the time window of the blink (Lag2 trials), but this seeming difference

was reduced outside of the blink window, for Lag8 trials (ON-M: 68.9%, SD: 14.9%; OFF-M: 67.8%, SD: 3.4%). However, this effect did not seem to depend on distractor valence, as DBS similarly affected target accuracy in negative vs. neutral trials: In Lag2 trials ON stimulation improved performance for neutral distractors by 8.5% on average (SD: 13.6%, range: -6% to +26%), and for negative distractors by 7.1% (SD: 6.3%, range: +1% to +15%). This difference was either smaller or absent for Lag8 trials: accuracy improved by 3% for neutral (SD: 14.8%, range: -13% to +19%), and decreased by 1% for negative distractors (SD: 10.2%, range: -16% to +7%). In line with our expectations, DBS thus seemed to selectively improve accuracy on Lag2 trials, but contrary to our expectations, the valence or saliency of the distractor mattered little.

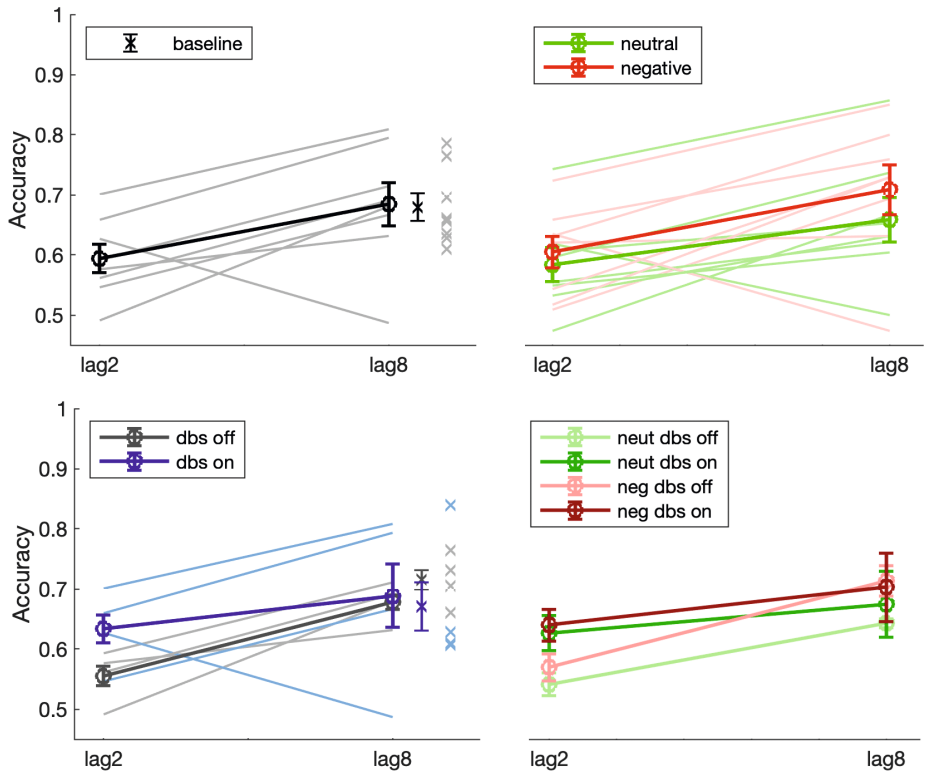


Figure 3. Mean accuracy for Lag across all eight sessions across four patients (top left), split by Valence (top right), DBS (bottom left), and their interaction (bottom right). Patients showcase a robust attentional blink on average, seemingly higher accuracy for negative distractors on Lag8 trials, and higher accuracy for DBS ON in Lag2 trials.

iEEG

We attempted to replicate and extend iEEG findings from Slagter et al (2017), where in a standard AB task, an increase in power in the α/β range (8-16 Hz) was present 80-140 ms after T1 (distractor in EIB) in T2 blink compared to no-blink trials, as well as an increase in the theta range (4-8 Hz) 150-400 ms after T2 was seen (target in EIB) in contrast with trials where it was unseen. Both expectations are marked in Fig. 4 with a black box. For neither the left or right striatal electrode no consistent effect can be seen after target presentation in the theta range in our data, neither OFF or ON DBS. There seems to be some consistency across patients for an effect in the α/β range on the left, but this effect was in the opposite direction from our prediction, and rather small in amplitude both for DBS OFF and ON, as can be seen in Figure 5. Thus, with our small number of cases, and a different blink paradigm, we did not reproduce the iEEG signatures of capture and conscious perception reported in Slagter et al. (2017).

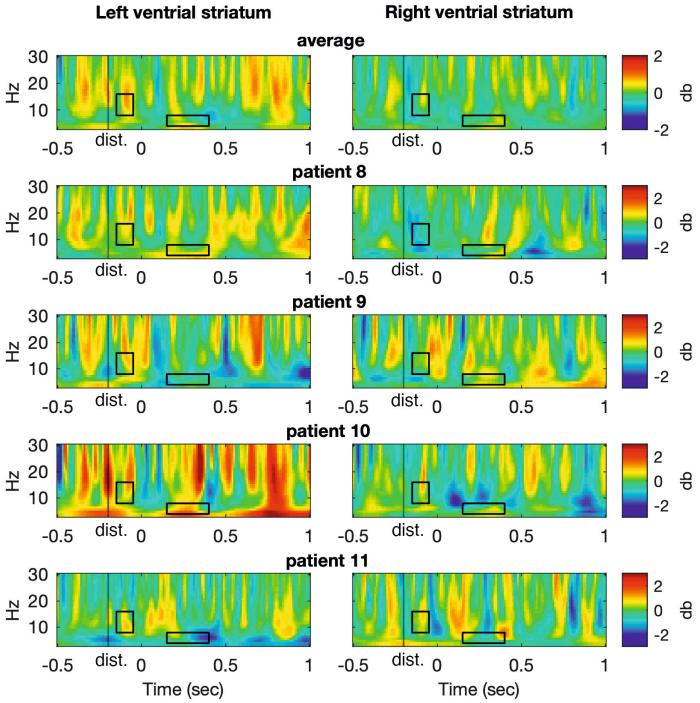


Figure 4. Time frequency results from the emotion-induced blindness task for seen versus unseen trials acrosss subjects (top row) and for each individual subject (lower rows). Presented are intracranial EEG data from the left (left figures) and right (right figures) ventral striatum. The target was presented at time zero. Time-frequency representations show in db the strength of differences in striatal activity between seen and unseen short-interval trials in the DBS OFF condition. The black rectangles denote the two time-frequency windows of interest.

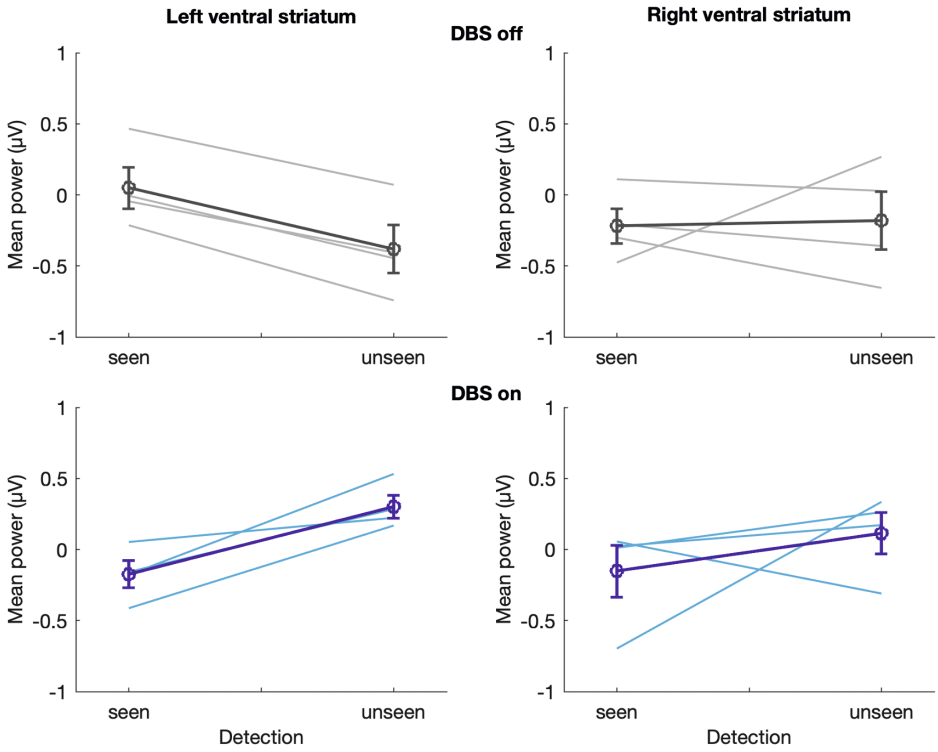


Figure 5. Presented is mean power for the predetermined α/β frequencies for the emotion-induced blindness task. As can be seen, we did not observe greater activity during DBS OFF in the α/β range to distractors in unseen vs. seen trials, in contrast to Slagter et al. (2017). For DBS ON as well, differences seem to deviate from zero only slightly.

Backward masking

Behavior

As can be seen from Figure 6, patient's accuracy improved with increased temporal asynchrony between target and mask (SOA), as is typically found (e.g., Breitmeyer, 2007). Whereas accuracy was at chance level for a SOA of 16 ms (M: 49.1%, SD: 5.8%), patients performed increasingly well with a SOA of 50 ms (M: 59.6%, SD: 10%) and even better at a SOA of 100 ms (M: 80.6%, SD: 17.4%). DBS ON improved accuracy for all three SOAs (16 ms M: +7.8%, SD: 5.9%, range: +2.5 to +14.2%; 50 ms M: +4.9%, SD: 8.4%, range: -1% to +14%; 100 ms M: +4%, SD: 10.9%, range: -2.2% to +16.6%). As this improvement was seen for all SOAs, and largest for the 16ms SOA, it is unclear if this effect reflects a true increase in accurate conscious perception or rather a response bias. This is unlikely

however given we found little change in confidence scores for low SOAs (see below).

We observed a similar pattern for response confidence: patients were increasingly confident in their response with higher SOAs: 16 ms (M: 36%, SD: 48.4%) compared to 50 ms (M: 62%, SD: 28%) and 100 ms (M: 90.1%, SD: 13.2%). DBS ON also numerically enhanced confidence scores increasingly so for longer SOA durations: for 16 ms (M: +.8%, SD: 4.2%, range: -2.2% to +5.6%), 50 ms (M: +2.3%, SD: 9.1%, range: -4.8% to +12.5%) or 100 ms SOA trials (M: +6.2%, SD: 10%, range: -1% to +17.7%).

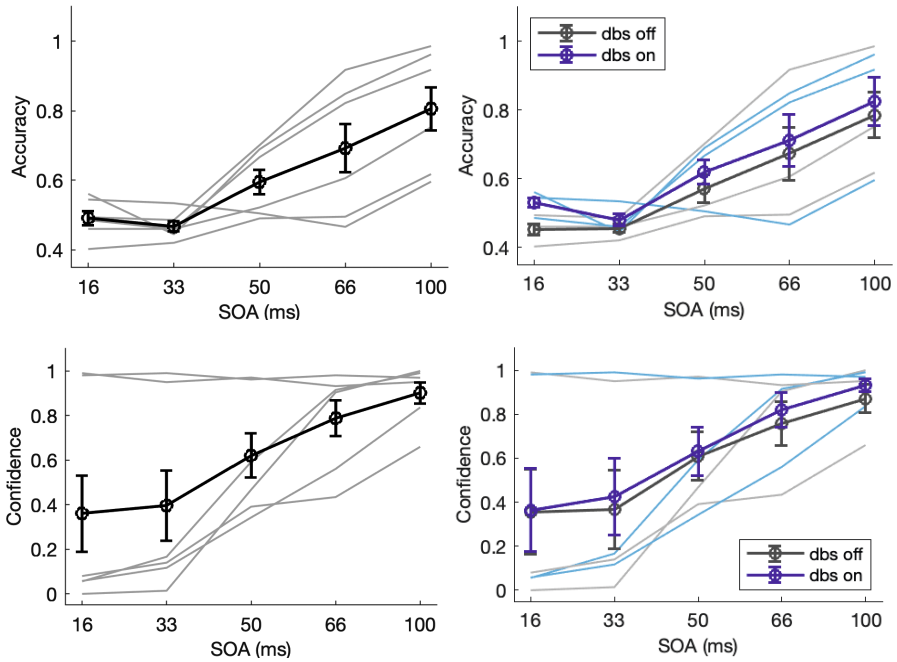


Figure 6. Mean accuracy and confidence, overall (left plots) and split by DBS conditions (right plots). Both accuracy and confidence scores improve with increased SOA (left plots; averaged across DBS OFF and ON). DBS generally enhanced accuracy, across SOA (top right plot), and confidence scores, in particular for the longer SOA durations (bottom right plot).

EEG

As shown in Fig. 7, we replicated previous reports showing that the target-evoked P3 ERP components are affected by the duration of target-mask SOA despite our small sample size (Boonstra et al., 2020; Del Cul et al., 2007). Specifically, the target-evoked P3 was virtually absent for the 16-ms SOA, in which targets were typically not seen, but present in all other SOA trials. While

the amplitude of the P3 did not further scale with longer target-mask SOAs, as in our previous study (Boonstra et al., 2020), P3 latency seemed to scale with SOA. This could potentially be explained by a more pronounced preceding N2 in longer SOA trials (see Fig. 7), which may have temporally overlapped with the P3. Past work has shown that the N2 is the first ERP component to linearly scale in amplitude with conscious visibility (Sergent et al., 2005). A roughly similar pattern of findings could be observed for DBS ON.

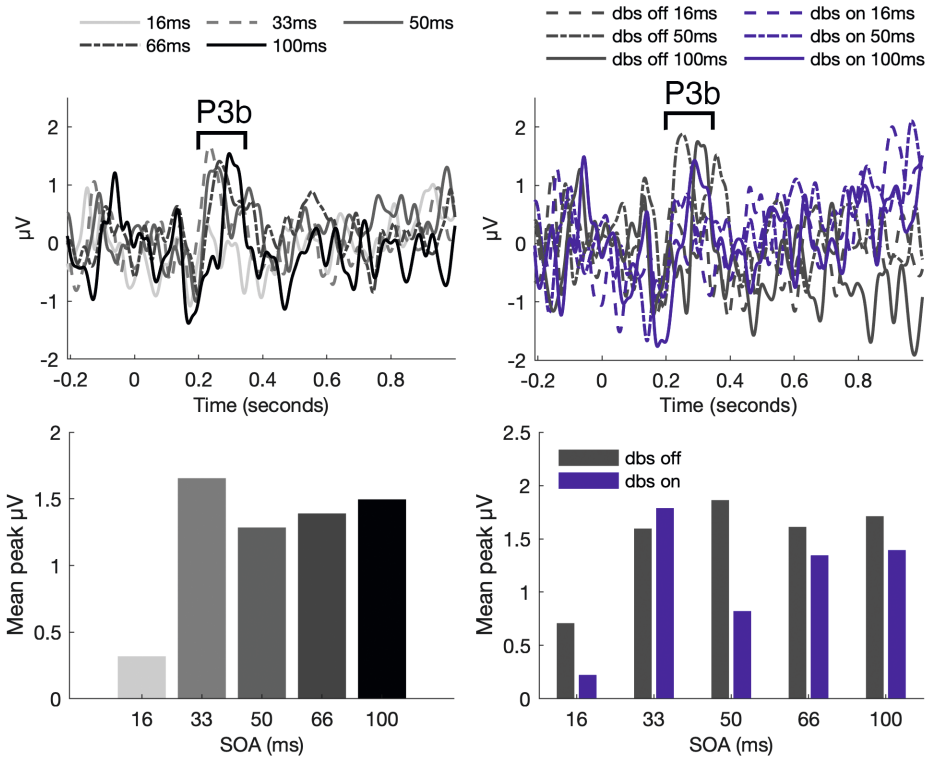


Figure 7. Effects of SOA and DBS on target-evoked ERPs in the backward masking task. This figure displays the grand-average target evoked ERPs for P3b electrodes (P1, Pz, P2, PO3, POz, PO4) for both DBS conditions combined (left panels) and separately (right panels), per target-mask SOA condition (16, 33, 50, 66, and 100 ms). This figure shows that ERP P3b amplitudes and latencies generally increased as a function of target-mask SOA, although not necessarily linearly (bottom figures).

iEEG

As in the case of the EIB task, we first tried to replicate the finding from Slagter et al (2017) concerning a selective theta response (4-8 Hz) between 150-400 ms to consciously perceived targets. As can be seen from Fig. 8, no such effect can

be shown consistently across participants. Due to a lack of the expected effect in the OFF condition, time-frequency spectra for the ON condition are omitted.

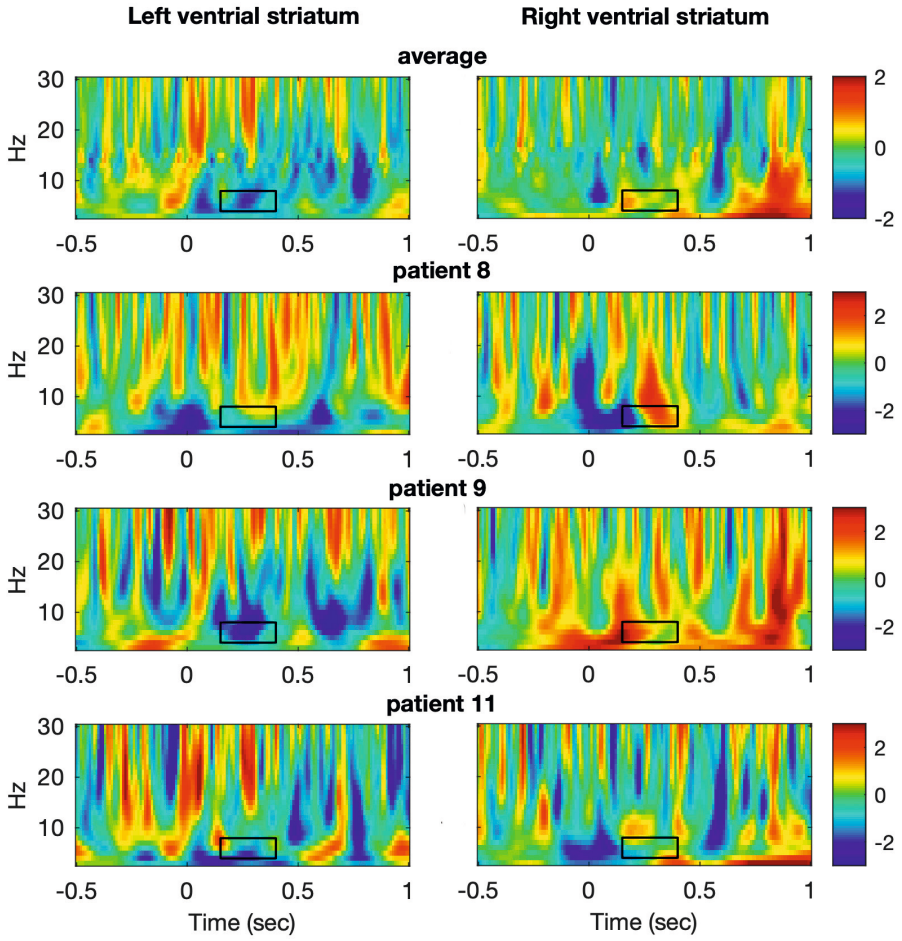


Figure 8. Time frequency results for seen versus unseen trials in the backward masking task OFF DBS. Intracranial EEG data are presented from the left (left figures) and right (right figures) ventral striatum. The target was presented at time zero. Time-frequency representations show in db the strength of differences in striatal activity between seen and unseen targets for 50 ms SOA trials with DBS OFF.

3.5. Discussion

In this preliminary study, we presented results showcasing how deep brain stimulation in the basal ganglia may influence conscious perception. As a

minimum requirement for the validity of our study, we showed how patients produced patterns of behavior expected for both emotion-induced blindness (EIB) and backward masking (BM). In addition, we showed that at a behavioral level there are indications suggesting that stimulation may improve target perception accuracy to a small degree in both tasks. Specifically, the emotion-induced blink was smaller ON DBS, and participants generally perceived more masked targets. Yet, our neural findings were less clear, conceivably due to our small sample size and their greater sensitivity to noise. We were able to largely replicate ERP findings from previous research for the BM task, showing a general increase in N2 and P3b amplitude and/or latency to consciously perceived targets as a function of SOA. Yet, we were unable to obtain reliable ERP responses to targets presented in an RSVP stream for the EIB task. Nor were we able to replicate time-frequency findings from previous research for either task OFF DBS. Lastly, we failed to show a meaningful effect of DBS on either ERPs or time-frequency decompositions. Thus, our findings suggest that ventral striatal DBS may improve conscious perception, but this critically awaits replication in large sample size studies that can also reveal the underlying neural mechanisms.

Currently, DBS is thought to affect neural tissue by generating efferent output at the stimulation frequency in the axon (Brocker & Grill, 2013; McIntyre et al., 2004). In the case of stimulation of the striatum in OCD patients specifically, the therapeutic effect of DBS seems to stem from the normalization of nucleus accumbens hypoactivity, and hyperconnectivity in the frontostriatal network (Acevedo et al., 2021; Figee et al., 2013; Smolders et al., 2013). More recently, two bundles of the anterior limb of the internal capsule have come to the forefront as target for stimulation: the anterior thalamic radiation (ATR) and the superolateral branch of the medial forebrain bundle (slMFB) (Coenen et al., 2012). The slMFB connects the prefrontal cortex via the nucleus accumbens to the ventral tegmental area (VTA), and contains dopaminergic projections from the VTA to the ventral striatum (Haber & McFarland, 1999). Improvements in mood are thought to contribute to the improvement of OCD symptoms through stimulation (Coenen et al., 2017). The ATR connects the prefrontal cortex to the anterior thalamus and is part of the cortico-striatal-thalamo-cortical network. Stimulation may normalize the dysregulation of this network in OCD (van den Heuvel et al., 2016). Stimulation of the slMFB seems most effective in combating OCD symptoms (Liebrand et al., 2019).

Given these courses of action, it is conceivable that the marginal behavioral improvements we recorded, stem from a normalized capacity for broadcasting of task-relevant stimuli, in line with theoretical accounts suggesting that

conscious perception depends information being made globally available (Crick & Koch, 2003; Dehaene & Changeux, 2011; Edelman, 2003). Indeed, an explanation of our behavioral findings in these terms would support the proposed role played by the BG in perceptual processes (Redgrave et al., 1999).

It should be noted however that several limitations pertain to our study. We tried to replicate findings from a previous iEEG study from our group (Slagter et al., 2017), showing how a distractor (T1) elicited α/early β response, possibly reflecting attentional capture, was associated with failed perception of a subsequent target (T2), as well as a theta burst only for seen targets (T2). A direct comparison between this study and the present study is impeded by the fact that both tasks weren't identical. While in the study by Slagter and colleagues (2017) a standard attentional blink task was used, in our study, patients performed an EIB task containing valenced distractors (equivalent to T1 in the AB task), and patients did not have to report the distractor. It must be emphasized there was good reason for this difference in design: we introduced a valenced distractor to strengthen the presumed attentional capture effect found previously with a neutral symbol. The idea was that if such a symbol was able to do so, then a negative image should elicit this capture effect to a greater degree, and we were interested to see if DBS could reduce this response.

Our sample size was reduced due to unforeseen limitations in DBS battery life, making it difficult to tell whether we succeeded in distilling signal from noise in our neural data. We also used a different stimulator than the study by Slagter and colleagues (2017). The unique advantage of this device is that it can stimulate and measure simultaneously, but a relative drawback is that the recording channel is online referenced to the top contact point on the same electrode through subtraction, which rendered it impossible for us to use the exact same referencing scheme as in Slagter al. (2017). Different contact points can show disparate and even opposite levels of activity, which complicates the comparison of our results with previous findings.

Despite these limitations, to our knowledge this is the first study to present data from patients performing an EIB and BM task ON and OFF DBS. We have shown there are indications to believe striatal DBS may improve conscious perception, but more research is needed to substantiate this claim.

Chapter 4

The dialectics of free energy minimization

Published as:

Boonstra, E. A., & Slagter, H. A. (2019). The dialectics of free energy minimization. *Frontiers in Systems Neuroscience*, *13*, 42. https://doi.org/10.3389/fnsys.2019.00042

4.1. Abstract

Karl Friston's free energy minimization has been received with great enthusiasm. With good reason: it not only makes the bold claim to a unifying theory of the brain, but it is presented as an a priori principle applicable to living systems in general. In this paper, we set out to show how the breadth of scope of Friston's framework converges with the dialectics of Georg Hegel. Through an appeal to the work of Catherine Malabou, we aim to demonstrate how Friston not only reinvigorates Hegelian dialectics from the perspective of neuroscience, but that the implicit alignment with Hegel necessitates a reading of free energy minimization from the perspective of Hegel's speculative philosophy. It is this reading that moves beyond the discussion between cognitivism and enactivism surrounding Friston's framework; beyond the question whether the organism is a secluded entity separated from its surroundings, or whether it is a dynamical system characterized by perpetual openness and mutual exchange. From a Hegelian perspective, it is the tension between both positions itself that is operative at the level of the organism; as a contradiction the organism sustains over the course of its life. Not only does the organism's secluded existence depend on a perpetual relation with its surroundings, but the condition for there to be such a relation is the existence of a secluded entity. We intend to show how this contradiction – tension internalized – is at the center of Friston's anticipatory organism; how it is this contradiction that grounds the perpetual process of free energy minimization.

4.2. Introduction

This paper moves in the interstices between neuroscience and contemporary readings of Georg Hegel's philosophy (Catherine Malabou, Slavoj Žižek). It seeks to show how Karl Friston's free energy minimization resuscitates Hegelian dialectics from the perspective of empirical science. Friston's approach to neural functioning has revolutionized our understanding of the brain. From the perspective of free energy minimization, the brain is no longer conceived as an organ that merely incorporates influences from outside; instead, the brain is viewed as an anticipatory structure engaged in the continual process of its own maintenance and transformation (Friston, 2010). This approach to neural functioning not only sits well with what we know about the brain's structure and functioning in terms of its anatomy and physiology, but the broader scope of Friston's framework – as well as the discussions surrounding it – beg for a return to Hegel.

The connection between Hegelian philosophy and neuroscience is not new. We find an emblematic example of their convergence in Malabou's book *What Should We Do with Our Brain?* from 2004, which revolves around the opposition between the Hegelian notion of plasticity on the one hand, and neuroplasticity on the other. Within neuroscience, we speak of neuroplasticity primarily in terms of the brain's capacity to undergo change (e.g., Buonomano & Merzenich, 1998; Feldman, 2009). For example, in terms of synaptic connections strengthening or weakening dependent on their excitation. This account of plasticity emphasizes *passivity*: the degree to which the brain changes in response to influences it undergoes from outside; it is the brain's capacity to *receive form*. Malabou shows how in Hegel's work, the notion of plasticity encompasses not only the designation of passively receiving form, but simultaneously implicates its obverse: the capacity to *produce* form (Malabou, 2005). Said differently, plasticity harbors activity that often gets lost in its neural variant. And yet, advances in our scientific understanding of neuroplasticity also touch on the activity of plasticity, by moving beyond the passive association of synapses. For example, 'homeostatic forms of neural plasticity regulate all the synapses on a neuron in unison in an orchestrated fashion.' (Sweatt, 2016: 189). At the network level as well, sensory and motor systems have the capacity (to some extent) for cortical remapping in response to disease or stroke, through which some functioning may be recovered in the aftermath of such events (e.g., Wittenberg, 2010). For example, '[e]vidence in sensory systems was obtained in work on crossmodal plasticity in which the loss of input to one sensory modality resulted in cortical reorganization in other sensory systems.' (Ostry & Gribble, 2016: 118). In both

of these examples, we are dealing with a form of neuroplasticity implicating an interplay between different levels of organization, where more general levels of organization (neuron, network) may subsume the behavior of more specific processes (synapses, brain areas). Once a process operates at a more general level of organization, it may no longer make sense to insist on reducing its functioning to more specific levels (Bello-Morales & Delgado-García, 2015). And yet, reorganization can occur bidirectionally: a change in a specific process (e.g., one sensory modality) may give rise to reorganization in general. This kind of self-relating organizational process is relevant for the present paper, because both in the case of Hegel and Friston, we are dealing with a bidirectional process of (active) organization and maintenance at the level of the brain or organism at large. The challenge at hand is to extend the brain's organizational capacity to our most basic understanding of neural functioning, in order to move beyond a conception of the brain as a passive receiver of influences.

The challenge we face in neuroscience is to provide a formal description of neural functioning *from the standpoint of the brain*. This means that our explanation should take account of the limited access the brain has to its surroundings, as well as the (computational) constraints the brain is subject to. As we will see over the course of this paper, Friston's free energy minimization is a formal description of neural functioning that adheres to these constraints. The second advantage of Friston's approach is that the framework is grounded in the minimal requirement for the existence and perpetuation of the brain itself: the imposition and maintenance of a boundary between the organism (including the brain) on the one hand, and the organism's surroundings on the other. The result is that the brain starts to appear not as an organ of infinite malleability and accommodation; not as a flexible organ where '[a]ll that is solid, melts into air' (Marx & Engels, 2002: 16). Instead, the brain appears as a self-organizing system that undergoes change as a way to *resist change*. Perhaps surprisingly, it is here where Hegel's 19th century philosophy and Friston's free energy minimization converge.

Indeed, perhaps it is the double meaning of plasticity – the capacity to receive form, and the capacity to produce form – which is at stake in free energy minimization. From this perspective, the brain is conceived as an anticipatory structure that actively models and anticipates its surroundings. An important implication is that both of plasticity's capacities are implicated in the same circular process; namely, in the form of the reception of sensation, and in the production of sensation through action. Under free energy minimization, the brain is caught in a perpetual process of self-production, on the basis of which form is not only received from outside, but is equally produced and maintained

by the anticipatory structure itself. Friston's framework allows us to see for the first time the significance of Malabou's invocation of the Hegelian notion of plasticity *from the standpoint of neuroscience itself*. The elucidation of Friston's free energy minimization against the backdrop of Malabou's reading of Hegel will allow us to mobilize Malabou's critique within a theoretical framework that has rapidly gained traction and stature within contemporary neuroscience.

There are two tenets from Malabou's reading that need to be brought to light in order to bring out the convergence between Hegel and Friston. First, the primacy of *habit,* understood as a reduplication of nature, through which the organism starts to appear, already in Hegel, as an anticipatory structure. Second, the processual nature of the formation and maintenance of habit; a process that is at once plastic and dialectical. What makes the process of plasticity dialectical is that in it, contradictory moments coincide. The changing *formation* of a particular organism simultaneously implicates its *deformation*. Framed in terms of old and new, the 'new defines itself in response to what is already established; at the same time, the established has to reconfigure itself in response to the new' (Fisher, 2009: 3). This contradictory process is what makes the organism (including the brain) dialectical.

But why appeal to Hegel in the first place? Why Hegelian dialectics? There are two reasons. The first reason is that already in his work we find the designation of living organisms in terms of differentiating processes that presuppose the imposition of a boundary. In both the case of Hegel and Friston, it makes sense to speak of organic life only on account of the existence of a boundary which constitutes and sustains a separation between inside and outside. Contemporary readings of Hegel's work suggest he was the first to fully anticipate contemporary conceptions of self-organization and self-production (*autopoiesis*). Due to the proximity in starting point and scope of both Hegel's and Friston's work, the first question we address is this: is free energy minimization a formalization of the dialectical process of plasticity, understood as the capacity both to receive and to produce form?

While Friston's approach to neural functioning is incredibly attractive, our wager is that Friston himself (Allen & Friston, 2018), as well as the appropriators of his framework in terms of cognitivism (Hohwy, 2013, 2016a; Wiese & Metzinger, 2017), and enactivism (Bruineberg et al., 2016; Clark, 2013, 2016b), do not do justice to the unsolvable tension at the heart of neural functioning under free energy minimization: the brain's anticipations are never "correct"; the brain necessarily and continuously sustains and attempts to solve the "errors" of its anticipations. There is no definite solution to this problem: the brain can

only optimize its anticipations and thereby minimize its error. The result is that under free energy minimization, the brain sustains tension necessarily.

In addition to the brain's unsolvable tension, the theoretical struggle surrounding Friston's framework revolves around the tension between cognitivism and enactivism; between seclusion and openness (Allen & Friston, 2018; Hohwy, 2016). This tension revolves around the question whether the brain is secluded from its surroundings, or whether it is better characterized by openness to these surroundings, as part of a larger brain-body-environment system. We are not appealing to Hegel merely for historical reasons; the second reason to appeal to his work is that his philosophy provides us with a way to move beyond this discussion. How? From a Hegelian perspective, it is not enough simply to choose either cognitivism or enactivism; seclusion or openness. Or to insist on finding the right balance between the two. The appeal to Hegel allows us to see how Friston's framework reintroduces the old Hegelian theme of a contradiction constitutive of life; an '*Unbehagen in der Natur*', or a discontent in nature (Žižek, 2016: 350-351). The challenge is to show how the unsolvable minimization problem of Friston's framework actualizes the tension between seclusion and openness. Said differently, we have to mobilize this tension at the level of the organism itself; to the point where it appears as a contradiction the organism sustains over the course of its life. The second question central to this paper is thus: can we enact the Hegelian shift from tension to contradiction with regards to free energy minimization?

The present paper consists of three parts. First, after outlining the contradictory organism in Hegel, and the dialectical/plastic process of habitual anticipation as understood by Malabou (Section 2), we will move on to Friston's conception of the anticipatory organism, in order to show how Friston's free energy minimization converges with Malabou's reading of Hegel (Section 3). Finally, we will try to move beyond both Friston and his cognitivist/enactivist appropriators, by enacting the shift from the organism's unsolvable tension, to its constitutive contradiction (Section 4).

4.3. Habitual anticipation, plasticity, dialectics

Science and teleology

The relationship between Hegelian dialectics and Friston's free energy minimization touches on a larger discussion surrounding the connections between self-organization, self-production (*autopoiesis*), Friston's framework, and predictive processing on the one hand, and Kant and Hegel on the other. The Kantian legacy at work in these notions, and in Friston's approach to the

brain, has received most attention within philosophy of mind (Hohwy, 2013; Swanson, 2018; Weber & Varela, 2002; Wiese & Metzinger, 2017). While some attention has been devoted to the relationship between dialectics in general and predictive processing (Bolis & Schilbach, 2018), an investigation into the specifically Hegelian legacy at work in free energy minimization is long overdue. Especially since the connections between Hegelian dialectics and the notions of self-organization and *autopoiesis* are well-established (Michelini, 2012; Michelini et al., 2018).

While Hegel can and should be criticized for relegating empirical science to something secondary (beneath philosophy), he by no means dismisses empirical science all together. On the contrary, his *Philosophy of Nature* is pervaded by the attempt to do justice to the science of his day: '[n]ot only must philosophy be in agreement with our empirical knowledge of Nature, but the *origin* and *formation* of the Philosophy of Nature presupposes and is conditioned by empirical physics.' (Hegel, 2004: 6, §246). However, Hegel lived long before the advent of non-equilibrium thermodynamics and cybernetics. The empirical physics in his time were Newtonian mechanics and the early days of thermodynamics. As such, Hegel followed Kant in problematizing the sufficiency of Newtonian mechanical laws for understanding living organisms (e.g., see Marques & Brito, 2014; Weber & Varela, 2002). In the words of Kant:

'An organized being is thus not a mere machine, for that has only a motive power, while the organized being possesses in itself a formative power, and indeed one that it communicates to the matter, which does not have it (it organizes the latter): thus it has a self-propagating formative power, which cannot be explained through the capacity for movement alone (that is, mechanism).' (Kant, 2000, 246, §65)

A central issue surrounding the 'self-propagating formative power' of organisms in contrast to Newtonian mechanics, is the problem of teleology, purpose, or "natural ends" [*Naturzweck*]. For Kant, a 'thing exists as a natural end *if it is cause and effect of itself*' (Kant, 2000, 243, §64). More precisely: '[a]n organized product of nature is that in which everything is an end and reciprocally a means' (Kant, 2000, 247, §66). In other words, the question of purposiveness in nature pertains to the way organisms exert purpose in the conservation and perpetuation of their own organization. Hegel upheld Kant's opposition between Newtonian mechanism and organism, as well as the problem of natural ends. However, when we speak of purpose, we need to distinguish between external and internal purposiveness:

'[t]he notion of end, however, is not merely external to Nature, as it is, for example, when I say that the wool of the sheep is there only to provide me with

clothes; for this often results in trivial reflections, [...], where God's wisdom is admired in that He provided cork-trees for bottle-stoppers, or herbs for curing disordered stomachs, and cinnabar for cosmetics.' (Hegel, 2004: 5, §245)

When Hegel speaks of internal purposiveness and the notion of end, he refers to a logic that pertains to the organization of natural objects:

'[t]he notion of end as immanent to natural objects is their simple determinateness, e.g. the seed of a plant, which contains the real possibility of all that is to exist in the tree, and thus, as a purposive activity is directed solely at self-preservation. [...]; the true teleological method – and this is the highest – consists, therefore, in the method of regarding Nature as free in her own peculiar vital activity.' (Hegel, 2004: 6, §245)

Purposive activity stems not from the projection of intentions, goals, or plans onto the organism. Instead, the organism's internal purposiveness is grounded in the immanent necessity of self-maintenance: internal purposiveness is self-preservation. The specifically Hegelian twist to the Kantian problem of "natural ends" resides in the assertion of purposiveness immanent to nature. For Kant, purpose remains ultimately at the level of knowledge and explanation; it is 'not an objective principle but a merely regulative one, a subjective maxim of the reflecting power of judgment. Therefore it has a value that is not constitutive but simply heuristic.' (Michelini, 2012: 135). For this reason, from the perspective of Kantian philosophy, we can only conceive of purposiveness "as if" it pertains to nature.

Incidentally, the opposition between external and internal purposiveness is central to Friston's free energy minimization as well. His framework provides an alternative to normative explanations 'in which I, as external observer, write down some condition for optimal behavior, rather than grounding that explanation in the necessary preconditions for the existence for that organism.' (Allen & Friston, 2018: 2473). By taking the preconditions of the organism as a starting point, 'the FEP [free energy principle] provides a normative, teleological essence to the synthesis of biology and information' (Allen & Friston, 2018: 2476). In other words, in the case of Friston as well, the organism's internal purposiveness ('teleological essence') follows from the necessary condition that has to be met for there to be an organism.

Both in the case of Friston and Hegel, such a necessary precondition is the existence of a boundary between the organism and its surroundings: 'the events that 'take place within the spatial boundary of a living organism' [Schrödinger] may arise from the very existence of a boundary or blanket, which itself is inevitable in a physically lawful world.' (Friston, 2013: 1). In Hegel's words: 'Nature's formations are determinate [*bestimmt*], bounded [*beschränkt*], and as

such enter into existence.' (Hegel, 2004: 284, §339). In other words, natural objects are to be bounded if they are to exist. As we will see in the next section, the gist of Friston's framework is that if an organism maintains a boundary, it minimizes free energy.

In the case of Hegel and his discussion of internal purposiveness, it is not enough to assert the importance of the existence of a boundary. If we were to leave it at that we would remain at a determination of the organism in terms of a bounded organization plus the magical power to exert purpose. There is no such "plus" in Hegel; he does not "add something" to mechanical nature; he does not succumb to vitalism. Instead of an addition, Hegel's solution to the problem of purposiveness consists in subtraction. As we will see below, it is what the organism lacks which gives it purpose or drive. By addressing what it lacks, the organism exerts purpose in the preservation of its organization. There is nothing mysterious about the notion of lack. If we regard an organism as an organized product of nature, then lack is simply a state of disorganization that needs to be addressed (e.g., hunger). Here too, there are striking parallels to Friston's framework in which the organism continually engages in minimizing the "error" inherent to its internal states, in an attempt to maintain an internal organization.

What makes the organism contradictory is the coincidence of organization and disorganization. 'For Hegel, life itself is imbued with contradiction because it is inextricably bound up with what it lacks: its identity is at one with its negation.' (Michelini, 2012: 133). Hegel's version of self-organization and self-production is thus a continuous process of self-preservation sustained by contradiction, through which a bounded organization is maintained. In the remainder of this section we will retrace Hegel's exposition, which results in the contradictory organism.[1]

1 There is much debate about what contradiction means in Hegel's philosophy (e.g., see Bole, 1987; Bordignon, 2017). While we may define contradiction as a pair of sentences, where one negates the other, it is important to emphasize that we are concerned not solely with epistemological contradiction. In Hegel's words, this means letting go of 'a tenderness for the things of this world' (Hegel, 2004: 5, §245), in order to assert the stain of contradiction at the level of things themselves. Over the course of this paper, we follow Žižek in the search for "contradiction' [inconsistency, impossibility, antagonism] as an internal condition of every identity.' (Žižek, 1989: xxix). In our case, this search pertains to the living organism under free energy minimization.

Animal organism

The entirety of Hegel's *Philosophy of Nature* forms a succession of stages that progresses from mechanics to organics, culminating in the animal organism. For our purposes, his discussion of the animal organism is particularly relevant, due to its striking similarity with Friston's free energy minimization. As such, we will focus on Hegel's discussion of the animal organism:

'The animal organism is the microcosm, the centre of Nature which has achieved an existence for itself in which the whole of inorganic Nature is recapitulated [*zusammengefa***ß**t] and idealized; this will be worked out in the detailed exposition to follow. Since the animal organism is the process of subjectivity, of self-relation in an outer world [*der Äußerlichkeit*], the rest of Nature is therefore here present for the first time as outward, since the animal preserves itself in this relation with the outer world [*zum Äußeren*].' (Hegel, 2004: 356, §352)

Almost everything we want to discuss in this section is contained in this passage. Let us unpack it. The advent of living beings for Hegel designates the moment where a natural process has extracted itself from the rest of nature and starts to function on its own terms. What gives animal life a privileged status is that the animal organism maintains a more determinate boundary with its surroundings compared to plant life: '[n]ow the plant is drawn towards the outer world but without truly preserving itself in connection with what is other, and consequently the rest of Nature is still not present for it as outer.' (Hegel, 2004: 356, §352). While a strong division between plant and animal life is problematic from the standpoint of modern biology (let alone the omission of other forms of life such as single-celled organisms), we are concerned with a simple point that Hegel places great emphasis on: plant life is not separated from its surroundings to the degree that animals are. 'The plant, as the first self-subsistent subject [...] still has its origin in immediacy' (Hegel, 2004: 304, §343). Put simply, this means that the plant cannot 'freely determine its place, i.e. *move from the spot*', and that 'its nutrition is not an interrupted process but a continuous flow' (Hegel, 2004: 305, §344). This changes with animal life, because 'the animal is a true, self-subsistent self which has attained to individuality, it excludes and separates itself from the universal substance of the earth which is for it an outer existence.' (Hegel, 2004: 355, §351). The animal maintains delineated break with its surroundings. While plant life 'still has its origin in immediacy', the animal organism has suspended the immediacy of its surroundings.

In the maintenance of itself as a distinct entity ('for itself'), the animal forms a recapitulation and "idealization" of inorganic nature. "Idealization" is what Malabou translates as "contraction"; around which Hegel's entire

Philosophy of Nature is organized: 'Hegel intends to show how the living organism summarizes everything that precedes it: inert matter, the elements, chemical processes, all the constitutive moments which are dialectically conjoined' (Malabou, 2005: 59). But the process of contraction goes further. As we will see further on, contraction designates the process through which the organism reproduces its constitution, in the broadest possible sense. The name for this constitution is habit.

In the suspension of immediacy; in the maintenance of a boundary, the animal organism has replaced natural immediacy with a second immediacy that is posited by the organism itself. This substitution is the 'characteristic of habit', through which nature is redoubled and as such starts to appear as 'second nature' (Malabou, 2005: 37-38). The living organism, and the animal in particular maintains a minimal difference between itself and its immediate surroundings. In the maintenance of this difference, the animal perpetuates itself as a habitual structure. The structural elements that constitute the animal organism – through which it contracts habit – are what Hegel calls 'Shape [*die Gestalt*]' and 'assimilation [*die Assimilation*]' (Hegel, 2004: 356, §352).

The animal organism's shape [*Gestalt*] is in turn comprised of three constitutive moments: 1) 'its simple, *universal being-within-self* in its externality [*allgemeines Insichsein in seiner Äußerlichkeit*]'; 'it is an undivided identity of the subject with itself – sensibility', 2) 'a capacity for being stimulated from outside and the subject's own reaction outwards to the stimulation – irritability', and 3) 'the unity of these moments, the *negative* return to itself from its relation with the outside world, and, through this, the production [*Erzeugung*] and positing [*Setzen*] of itself as a singular – reproduction, which is the reality and the basis of the first two moments.' (Hegel, 2004: 357, §353; see Michelini, 2012; Michelini et al., 2018, Fig. 1, for a full schematic overview of the animal organism in Hegel's *Philosophy of Nature*).

As is characteristic of Hegel's philosophical exposition, these three moments can be regarded separately, in which case they remain abstract, but only when considered together as a single movement do they do justice to the totality of the process constituting the animal organism:

'Reproduction passes through sensibility and irritability and absorbs them; it is thus derived, posited universality which, however, as self-producing [*das Sichproduzieren*], is at the same time concrete singularity. It is reproduction which is first the whole – the immediate unity-with-self in which the whole has at the same time entered into relationship with itself.' (Hegel, 2004: 358, §353)

In other words, reproduction here does not refer to replication, but to 'self-production, or the active conservation of a self-produced identity' (Marques

& Brito, 2014: 92). The totality of the movement of reproduction includes sensibility and irritability. Let us start with sensibility, which designates the identity of the animal organism with itself: the 'the sentient creature [*das Empfindende*]' (Hegel, 2004: 358, §354). When considered in isolation, the 'system of sensibility' is constituted by 'the extreme of *abstract* self-reference' (Hegel, 2004: 357, §354), which simply designates the rudimentary differentiation of the animal with regards to its immediate surroundings, whereby it maintains a minimum of autonomy.

However, the animal not only exists in isolation: the system of sensibility is differentiated outwards through the nervous system, on account of which the animal has 'an inward and outward reference – the sensory and motor nerves respectively' (Hegel, 2004: 359, §354). Indeed, it is through the nervous system that the animal is sensitive to (outside) influences:

'The moment of difference in sensibility is the nervous system which is directed outwards and is involved in external relationships: sensation [*Empfindung*] as determinate – either as immediately posited from outside or as self-determination [*Fühlen oder Selbstbestimmung*]. The motor nerves mostly start from the spinal cord, and the sensory nerves from the brain: the former are the nervous system in its practical function, the latter are that system as receptive of determinations, and to this the sensory organs belong.' (Hegel, 2004: 363, §354)

The first thing to note in this passage is that Hegel introduces a rudimentary circular process ('self-determination') through which the organism can determine its own sensations on the basis of its motions. In the next section, we will see that such circularity is crucial within Friston's free energy minimization. The second thing to note is that, as the animal is 'receptive to determinations', it not only maintains a break with its surroundings, but it also maintains a break with itself. As the animal 'enters into relationship with itself', the animal 'has itself for its object' (Hegel, 2004: 353, §351). This self-relation, self-differentiation, or self-feeling is sensation, which is the 'absolutely characteristic feature [*Bestimmung*] of the animal.' (Ibid.). As we will see later on, free energy minimization also functions on the basis of the incorporation of sensation or differentiation, in the form of anticipatory "error".

The movement of self-determination not only pertains to differentiation inwards; the animal also differentiates itself outwards: 'the subjectivity of the animal is not simply distinguished from external Nature, but the animal distinguishes itself from it; and this is an extremely important distinction, this positing of itself [*Sichsetzen*] as the pure negativity of *this* place, and *this* place, and so on.' (Hegel, 2004: 354, §351). Which brings us to the second moment:

'irritability is just as much a capacity for being stimulated by an other and the reaction of self-maintenance against it, as it is also, conversely, an active maintenance of self' (Hegel, 2004: 359, §354). In other words, building on top of the capacity for self-differentiation or sensation, irritability designates the organism's capacity to react to, or act on, its sensations. In the combination of the self-determining loop of sensibility, and the capacity to react and maintain itself in irritability; we have the necessary ingredients to complete the overarching loop that constitutes the third moment: reproduction. In passing through both sensibility and irritability, the animal continually (re-)produces and maintains itself as a distinct entity. Let us continue to how the totality of this movement operates in the next stage Hegel distinguishes.

The organism's reproductive movement actualizes in relation to its surroundings, through which the organism appropriates these surroundings:

'Now since the organism is directed towards the outer world as well as being inwardly in a state of tension towards it [*innerlich dagegen spannt*], we have the contradiction of a relationship in which the outer must be sublated [*aufgehoben*]. The organism must therefore posit what is external as subjective [*das Äußerliche als subjektiv setzen*], appropriate it, and identify it with itself; and this is assimilation.' (Hegel, 2004: 381, §357)

In other words, as the animal necessarily stands in relation to its surroundings; its reproduction takes the form of perpetual activity, through which the animal assimilates nature external to itself. The animal's relation to food makes up the most elementary example of this process. As the organism strives to overcome the deficiency of hunger, it passes through the three stages of sensibility, irritability, and reproduction in the corresponding sequence of hunger, ingestion, and digestion (Michelini, 2012; Michelini et al., 2018: 8). As such, the organism reproduces itself through the ingestion and subsequent digestion of food.

The animal's capacity for sensation permits the experience of tension: the 'self-feeling of the individuality is also directly exclusive and in a state of tension with a non-organic nature which stands over against it as its *external* condition and material.' (Hegel, 2004: 380, §357). And while this material may be employed to alleviate states of lack or deficiency such as hunger, the overcoming of deficiency offers only temporary respite: 'the animal perpetually returns from its satisfaction to a state of need.' (Hegel, 2004: 391, §362). In other words, the animal only temporarily overcomes its deficiency and necessarily revisits and maintains a state of tension with its surroundings.

While hunger is the most readily available example of this process, the simplicity of the example should not deceive us: its basic logic is 'particularized

in an infinite variety of ways' (Hegel, 2004: 388, §360). While the example of hunger is intuitive, it does not go far enough, because in order to speak of plasticity in terms of simultaneous reception and production of form, we need a process that not only alleviates states of need, but we need a process able to reconfigure the organism's habitual structure itself. Indeed, the generalized logic of "digestion" is what we are dealing with in the process of assimilation, where the animal contracts and posits its surroundings as part of its own habitual structure. The result of this continuous abstraction-contraction-recapitulation is 'literally, habitus, at once the internal disposition and the general constitution of the organism' (Malabou, 2005: 59). The "general constitution" of the organism refers to the recapitulation of inorganic nature that came before it, while the organism's "internal disposition" signifies the organism's contracted habitual structure. As we will see later on, it is the generalized logic of perpetual contraction that we find in Friston's free energy minimization, in the sense that the configuration of an anticipatory model "feeds off" its surroundings.

Contradiction and anticipation

We began with the organism understood as a self-enclosed entity; an abstract 'system of sensibility' which is subsequently differentiated outwards on account of the nervous system. In the second moment of irritability, the organism is not only receptive to stimulation, but reacts to such stimulation by engaging in active self-maintenance. Through these moments of sensibility and irritability, the organism embodies a self-productive loop in which its surroundings are assimilated. With these two stages of shape [*Gestalt*] and assimilation, we passed from an abstract to a relational understanding of the organism. We now have the necessary ingredients to take the third and final step in order to show what makes the organism contradictory.

If we emphasize the organism's autonomy, the problem is that the relation of the organism to what lies outside it remains an external relation. Its conception remains abstract; as if the organism has a choice to engage or not with its surroundings. If we take a step further, we are forced to admit that the organism "always already" stands in relation to its surroundings, as it is internally strung in opposition to it. However, if we emphasize the organism's perpetual relationship to its surroundings, we risk losing the very condition for its existence as a distinct entity: the boundary between organism and surroundings. The difficulty resides in conceiving of the organism as both abstract and relational. More precisely, the point is not to choose either an abstract or a relational conception of the organism. Rather, the point is to

see how the organism's relation to and dependence on its surroundings is simultaneously constitutive for the organism understood as a distinct or an "abstract" entity. Therein resides the organism's contradiction:

'Although common thought has it that need indicates dependence on something else, in reality, in a paradoxical way, it is a manifestation of independence: in fact water and food would be totally indifferent to the living being and they would not be able to have a 'positive' relation with it if the living being was not, for Hegel, 'the possibility of this relation'.' (Michelini, 2012: 137)

Said differently, the paradox is that the organism's relation to something outside itself is simultaneously the guarantor for the minimum of its autonomy. There is no choice to be made here between abstract and relational: the organism itself is subject to the tension between independence and dependence in the perpetuation of its life. For there to be a relationship of dependence between organism and surroundings, there needs to be an independent organism in the first place, in the sense of a distinct entity that is able to engage in a relationship of dependence. We can also turn this around: the only way for the organism to conserve and perpetuate itself as an independent entity, is to engage in a continuous relationship of dependence with its surroundings, through which the organism assimilates external nature. This relationship of dependence in turn is continually reinvigorated by the organism's recurring state of need. In this sense, the tension between independence and dependence is operative at the level of the organism itself. Insofar as the organism maintains a boundary between itself and its surroundings, it perpetually revisits a state of tension in need of alleviation. As such, sustained tension is concomitant with the existence of a boundary. As we will see below, this tension internal to the organism's organization is the sustained contradiction which serves as a precondition for the organism's existence concurrent with the boundary that constitutes its life.

With the organism's constitutive contradiction, we return to the problem of internal purposiveness from the start of this section. For Hegel, the organism's loop of self-determination is sustained by the contradiction it sustains, which gives the organism its internal purposiveness: '[n]eed and drive are the readiest examples of [internal] purpose. They are the felt contradiction, as it occurs within the living subject itself; and they lead into the activity of negating this negation [...].' (Hegel, 2004: 281, §204). The living organism (subject) feels contradiction, and its activity is directed towards overcoming this contradiction by getting rid of (negating) the feeling of need: the process begins with 'the feeling of *lack* [*Mangel*], and the urge [*Trieb*] to get rid of it [*ihn aufzuheben*]' (Hegel, 2004: 385, §359). But because the organism's contradiction is constitutive for its existence, there is no definite escape from it as long as the organism

is alive: '[t]his contradiction, that they are and are not, [...], manifests itself as a perpetual process' (Hegel, 2004: 6377, §356). In its perpetual attempts at overcoming its needs, the animal engages in 'activity of deficiency' (Michelini, 2012: 137).

Therein resides the paradoxical status of living beings: the organism harbors at once the maintenance of its own identity, as well as the negation of this identity. 'The defect [*Mangel*] in a chair which has only three legs is in us' (Hegel, 2004: 387, §359). The organism functions like a three-legged chair that stays upright, but continually slants into the direction of its missing fourth leg. As Francisco Varela put it: 'If we invert our perspective, this constant bringing forth of signification is what we may describe as a permanent lack in the living: it is constantly bringing forth a signification that is missing, not pregiven or pre-existent.' (Varela, 1997). In the continuous attempts at suspending its needs, the organism reproduces itself by contracting its surroundings. It is in this sense that the organism organizes itself around the contradiction that constitutes its existence. Insofar as this organization endures, the name for this organization is habit. As we will see later on (Section 4.3), in mobilizing this contradiction immanent to the organism, Hegel anticipated contemporary discussions surrounding the appropriation of Friston's framework in terms of cognitivism and enactivism.

What makes Hegel's philosophy difficult to follow is that he forces us to retrace numerous processes separately that he subsequently ties in together. What makes it outright frustrating is that the tension within the exposition itself is transposed into the object under study. The result is that by the end of the exposition, our initial starting point has been problematized. This is what Helmholtz missed when he dismissed Hegel for starting with the 'hypothesis of Identity' (Helmholtz, 1995: 79). In our case, the abstract conception of the animal organism as a self-enclosed entity is transformed through a relational designation, into an entity that appears as inherently contradictory. In this sense, the same circular structure of the organism's self-production also pertains to Hegel's philosophy itself. His philosophical exposition takes the form of a circular process, which, at the end of the circle, has retroactively undermined its own starting point. Again, in the case of the organism, the point is not to exchange an abstract conception for a relational one, from where the organism is engaged in continuous exchange with its surroundings. The organism is both abstracted from its surroundings as it maintains its boundary, *and* it stands in constant relation to these surroundings. The point is to recognize that the tension between abstract and relational is actualized and operative at the level of the animal organism itself in the form of recurring lack.

It is this contradiction between habitual structure and lack, or organization and disorganization, which gives rise to the organism's internal purpose or drive.

Hegel's animal organism thereby goes beyond (Newtonian) mechanism under immediate influence of external nature: the 'organism is no longer "determined by external causes" but irritated by external forces.' (Marques, 2016: 128). The suspension of the immediacy of external nature functions on the basis of the boundary that separates the organism's "second nature" or habit. It is on account of the animal's contracted habitual structure and the deficiencies that mark this structure, that the animal operates on the basis of its own purposiveness: it 'is only as this self-reproductive being, [...], that the living creature *is* and *preserves itself*; it only is, in making itself what it is, and is the antecedent End which is itself only result.' (Hegel, 2004: 356, §352). How does the contraction of habit relate to the notions of plasticity and dialectics?

Both notions designate not a property or attribute of the living organism, but they are two ways of approaching the living organism *as a process*. This is crucial, because for Hegel, '[s]tructure, as alive, is essentially process' (Hegel, 2004: 377, §356). As we briefly stated in the introduction, plasticity in Hegel designates 'a capacity to receive form and a capacity to produce form' (Malabou, 2005: 9). The process we have been describing so far is plastic: in the contraction of its surroundings, the organism is not only formed by its surroundings, but it produces its own form in the process. The moment forces external to the organism cross over onto the terrain of the organism, the organism posits and molds external nature as part of itself.

The process of plasticity is simultaneously dialectical, because the operations which constitute it, 'the seizure of form and the annihilation of all form, emergence and explosion, are contradictory' (Malabou, 2005: 12).[2] The formation of habit does not pertain to a stable entity to which habits are added and subtracted like attributes. The formation of habit *is the formation of the organism as such:* '[h]abit is there not only as a particular momentary satisfaction; rather I am this habit. It is my universal mode of being—what I am is the totality of my habits. I can do nothing else, I am this.' (Hegel, 2007: 153).

2 In classical logic, the principle of explosion is often expressed as *ex contradictione quodlibet sequitur* [from contradiction, anything follows] (e.g., see Lopez-Astorga, 2016). In Malabou's work, "explosion" pertains to plasticity as 'the capacity to annihilate the very form it is able to receive or create.' For example, the French '*plastique*, [...], is an explosive substance' (Malabou, 2008: 5). In this sense, while plasticity enables a structure of anticipation, it simultaneously upholds the possibility of this structure to deform itself, as well as radical openness towards the future.

It is because habit constitutes the organism that every change to its formation implies a simultaneous deformation. This coincidence of contradictory moments is what makes the process of plasticity, the contraction of habit and the perpetual reproduction of the organism, dialectical.

It is also the process of plasticity which makes the animal anticipatory: '[n]eed, appetite, desire, the accumulation of such retentions and expectations, are themselves proof of the fact that the animal is concerned to ensure the perpetuation of its own life' (Malabou, 2005: 64). In the reproduction of itself, the animal not only maintains a relation to itself and its surroundings, but it also stands in relation to its future: 'it is the structure of anticipation through which subjectivity projects itself in advance of itself, and thereby participates in the process of its own determination' (Malabou, 2005: 18). The perpetual reappearance of need and the drive to overcome need, indicate the animal's anticipatory disposition. More precisely, they indicate that the animal itself is simultaneously a rudimentary *structure and process* of anticipation. In the perpetual process of its own restructuring, the animal posits the presuppositions of its own anticipations. Hegel and Friston share the same starting point: a boundary between organism and surroundings, but can we conceive of Friston's anticipatory brain as a process that is at once plastic and dialectical?

4.4. States of anticipation

Anticipatory brain

The breakthrough of Friston's free energy minimization stems from the unconventional answer it provides to a simple question: what does the brain do? From the perspective of Friston's framework, neural functioning is subordinated to the overall imperative of an organism to maintain a boundary (Allen & Friston, 2018; Ramstead et al., 2018). In the process of boundary-maintenance, the brain is caught in a continuous process of anticipation in terms of both perception and action. It is this approach to the brain that has been well received in philosophy of mind (Bruineberg et al., 2016; Clark, 2013, 2016b; Hohwy, 2013, 2016a; Wiese & Metzinger, 2017). The conception of the brain as an anticipatory or predictive structure forms the first substantial challenge within neuroscience to the conception of the brain in terms of the "computer metaphor". In the traditional view, we approach the brain as a computer that primarily processes information from outside; similar to the way data is fed into a computer, which is subsequently processed and potentially retained for later use. While such bottom-up, stimulus-driven processing undoubtedly makes up

an important part of what the brain does, anticipatory accounts suggest that the importance of such bottom-up processing has been overstated. From the perspective of these accounts, a (matured) brain is engaged to a much greater degree in the prediction or anticipation of its input. Instead of processing influences from outside, the brain first and foremost constitutes its reality on the basis of what it learned to expect, onto which influences from outside intrude.

It is not the case that the anticipatory perspective introduces top-down influences into the way we view the brain. The traditional view also assigned importance to top-down influences in the form of memory, cognitive control, attention etc. In the traditional view however, such top-down processes remained secondary in dealing with influences from outside. From the anticipatory perspective, external influences become secondary to anticipatory states. What is so radical about this conception is that it turns the traditional view of neural functioning on its head: external influences get caught up in the primary process of the brain's anticipatory activity. As we will see below, the brain's main locus of activity becomes the minimization of the errors stemming from its own anticipations, through which it iteratively settles on anticipations of the future. The brain's imperative becomes to explain away the mismatch (prediction error) between its anticipations (top-down) and its actual input (bottom-up), across the cortical hierarchy. This mismatch in turn serves as a driver for change in the structure that generates future anticipations. It is important to emphasize that such anticipations do not pertain to perception alone. The perceptual aspect of anticipation stands in service of action. In addition, the input that potentially perturbs the brain's anticipations is not only passively received from outside, but can equally be elicited by the structure's own actions (Adams et al., 2013). That is, through action, the brain can actively change its outer world into a state the brain itself anticipated; minimizing prediction error in the process.

An important advantage of Friston's framework is that it stays close to the anatomical and physiological organization of the brain. For example, the primacy of top-down predictions over bottom-up input fits well with findings showing that top-down connections are both more divergent and abundant than bottom-up connections (Friston, 2005). In addition, the framework is able to adhere to the local computational constraints that individual neurons are subject to (Bogacz, 2017). The criticism that free energy minimization is unfalsifiable misses the point: Friston's framework does not stand in contrast to traditional neuroscience at the level of testable theories; it is opposed to

traditional neuroscience precisely at the level where *traditional neuroscience itself is unfalsifiable.* As in the case of the computer metaphor.

The opposition between anticipatory accounts and traditional neuroscience touches on two problems discussed earlier: how to do justice to the active organizational capacity that pertains to the brain's plasticity? And how to conceive of the 'self-propagating formative power' that pertains to living organisms? In answering these questions, a simple action-reaction (or stimulus-response) account seems inadequate, but we simultaneously need to avoid adding external purpose to mechanical nature. The often-employed way out of this deadlock in biology is to approach living organisms as dynamical systems (Ashby, 1966; Bertalanffy, 1950; Lewontin & Levins, 2007; Maturana & Varela, 2012). From the perspective of these approaches, living organisms are conceived as systems that maintain themselves through a process of self-organization and self-production (*autopoiesis*). It is such a process that Friston formalizes with free energy minimization, which attempts to capture brain function in the broadest possible sense. In the remainder of this section, we will attempt to elucidate Friston's framework (Section 4.2), in order to bring out its convergence with Malabou's reading of Hegel (Section 4.3).

Free energy minimization

Free energy minimization is so broad in fact, that it does not pertain solely to brain function; it is a framework pertaining to living systems as such. For this reason, Friston not only reopens the question "what does the brain do?", but the same goes for the even broader question: what is life? The specific organization of the brain is not primary, but rather 'the consequence of, or requirement for, this fundamental imperative [of free energy minimization]' (Friston, 2013a). As such, we will speak primarily of the organism from here on out, which subsumes the anticipatory brain. Friston suggests

'that biological self-organization is not as remarkable as one might think—and is (almost) inevitable, given local interactions between the states of coupled dynamical systems. In brief, the events that 'take place within the spatial boundary of a living organism' may arise from the very existence of a boundary or blanket, which itself is inevitable in a physically lawful world.' (Friston, 2013b)

In order to make sense of Friston's framework, we have to adopt the language of statistical physics. This means that we approach the organism as a system that stands in relation to its local surroundings or environment. Because the system is always already embedded in this relation, it necessarily lacks a complete overview of both its own possible states, as well as the possible

states of its local surroundings. Free energy minimization comes down to a computational problem in which the organism attempts to minimize the mismatch between its internal states on the one hand and the (inferred) external states of its surroundings on the other. What is a state? 'Formally speaking, the state of a system corresponds to its coordinates in the space of possible states, with different axes for different variables [...].' (Friston, 2018a). Put simply, a multitude of states forms the structural configuration of the organism: 'the repertoire of physiological and sensory states in which an organism can be is limited, and these states define the organism's phenotype' (Friston, 2010: 127).

With the notion of "state" we touch on a common thread with regards to Hegel. Hegel's notion of "habit" stems from his reading of Aristotle's *De Anima*. The etymology of the word 'habit' leads back to the Latin *habere*: "a way of having", and the Greek verb ἔχειν. 'This verb means 'to have', but as soon as it is followed by an adverb, it changes its meaning to include 'the *state* of being in one way or another' (Malabou, 2005: 37; italics ours). We thus find here an additional indication that both lines of thinking are closer than they may appear: both "state" and "habit" refer to the configuration of the anticipatory organism.

The internal states of the system and the external states of its surroundings are separated by the organism's intermediary layer comprised of sensory and action states. For the organism, the problem of maintaining itself consists in coordinating the (dis)concordance between different kinds of states. From a perspective outside the system, the difficulty in understanding how the coordination of these states unfolds, resides in that we are dealing with a circular structure, where each kind of state presupposes another. A sensory state presupposes the evocation of said state, either from outside, or by the organism itself. If the organism acted on its own sensory states, such action presupposes an anticipation on the basis of which to act. Such an anticipation in turn require internal states on the basis of which to anticipate. And finally, the existence of an internal organization necessitates the existence of a delineated organism in the first place. We are thus dealing with a loop in which internal states, action states, external states, and sensory states mutually implicate each other. In other words, the existence of a boundary between organism and surroundings is a necessary precondition for the process to take place; the boundary itself 'induces a circular causality' (Allen & Friston, 2018: 2474).

The proximity to Hegel is clear. In his work as well, the discussion of the animal organism presupposed differentiation: a bounded organization that sustains a break with external nature. Insofar as the animal organism is alive it upholds a minimum of autonomy by maintaining a boundary between itself and its surroundings. In addition, animal life for Hegel is also caught in a

loop of sensation and motion, which gives the organism the capacity for self-determination. In the case of Friston, the idea is that if a separation between internal and external states exists, then the system must minimize free energy. Why?

The logic is as follows: out of all possible states the system can be In, there is a relatively small number of states that sustain its life. This implies that, if an organism endures, it will necessarily frequent a limited number of states that lie within its physiological bounds. The maintenance of a limited number of states implies that the system will need to try to minimize surprise with regards to its own states. For example, ''a fish out of water' would be in a surprising state (both emotionally and mathematically).' (Friston, 2010: 127). Under the assumption that the system limits its own states, 'the long-term average of surprise is entropy. This means that minimizing free energy – through selectively sampling sensory input – places an upper bound on the entropy or dispersion of sensory states' (Friston, 2012: 2). The maintenance of such a limited number of states designates the way biological systems resist 'the dispersive effects of fluctuations in their environment' (Friston, 2012: 1). In other words, the minimization of free energy designates how an organism resists its surroundings and thereby maintains an internal organization.

The placement of an upper bound on surprise simultaneously implies its obverse; namely, 'systems that minimize free energy also maximize a lower bound on the evidence for an implicit model of how their sensory samples were generated' (Friston, 2012: 2). This 'implicit model' is the anticipatory structure generating the organism's anticipations. Such a generative model 'aims to capture the statistical structure of some set of observed inputs by tracking (one might say, by schematically recapitulating) the causal matrix responsible for that very structure' (Clark, 2013: 2). Taken together, free energy minimization implies that the system limits the states it frequents (by placing an upper bound on the surprise of these states), and thereby maximizes the evidence for the 'implicit model' that it is. As in the case of Hegel, this model is constituted in a continuous process of contracting (schematically recapitulating) the organism's surroundings.

But what is free energy? What does the organism minimize exactly? Friston says:

> 'Free-energy is an information theory quantity that bounds the evidence for a model of data. Here, the data are sensory inputs and the model is encoded by the brain. More precisely, free-energy is greater than the negative log-evidence or 'surprise' in sensory data, given a model of how

> they were generated. Crucially, unlike surprise itself, free-energy can be evaluated because it is a function of sensory data and brain states. In fact, under simplifying assumptions [...], it is just the amount of prediction error.' (Friston, 2009: 293)

What sets Friston's framework apart from just another formal description of the organism's behavior is that it formalizes how the organism (including the brain) is able to maintain itself *from the perspective of the system itself.* This starting point is crucial, because it takes serious how the system lacks access to the distribution of all possible states external to it. It cannot know all possible configurations of its surroundings. As a matter of fact, the system does not have access to external states at all, apart from the sensory states these external states elicit. In this sense external states are "hidden". It is because external states and their distribution are only indirectly accessible that the organism is unable to minimize surprise directly. Instead, the organism engages in the minimization of an upper bound on surprise, which is a problem the organism can solve, because it relies not on having access to the distribution of possible external states, but only on internal states (the model in the citation above) and elicited sensory states (the data). The system minimizes this upper bound on surprise either by changing its sensory input through action; a process Friston calls active inference; or by altering its internal anticipatory model: perceptual inference. On the one hand, 'action is the only way to underwrite an upper bound on the entropy of sensations. On the other hand, perceptual inference is the only way to inform action.' (Allen & Friston, 2018: 2477). For example, if the system cannot make out what caused its sensations, in addition to inferring a probable percept given the context, another option is to move closer in order to get a better look. In other words, the system has two ways of minimizing free energy; it minimizes the error of its own anticipations through perceptual and active inference.

In the process of inference, anticipatory or prediction "errors" refer to the deviation of sensory data from the predictions or anticipations formulated by the system. For this reason, it is important to emphasize that these "errors" are "wrong" only in relation to the anticipations that arise from the system itself. The standard by which an "error" is determined is strictly immanent to the system. How the system deals with prediction error is important, because these errors determine both the system's configuration, as well as the actions the system engages in. On the one hand, if these errors are not taken serious enough, the system is at risk of detaching itself from the demands of its surroundings. On the other hand, if these errors are taken too seriously, the system may find

itself engrossed by every insignificant change in scenery. In order to cope with anticipatory error, the system's challenge is not only to infer its external states, but the system also needs to determine the degree of certainty or precision of its inferences. The perspective free energy minimization subscribes to 'suggests that there are only two sorts of things that need to be inferred about the world; namely, the state of the world and uncertainty about that state.' (Friston et al., 2012: 2). To connect with what we said above: uncertainty is another way to conceive of entropy; to minimize free energy means to resist being overwhelmed by uncertainty: '[o]ur objective, given a model (brain), *m*, is to minimize the average uncertainty (entropy) about some generalized sensory states' (Feldman & Friston, 2010: 5). In order to constrain the uncertainty of its states, the system needs to estimate the precision of the disruption (prediction error) of its anticipatory states. If such disruption is estimated to be precise, it will exert more influence on future anticipations compared to imprecise disruption. In the interplay between anticipation and disruption, the dominance of either is dependent on the precision afforded to both. This enduring conflict amounts to how something is perceived and what is to be done.

An important implication of Friston's framework is that it gives a purely formal description of what the system does in terms of boundary preservation. In doing so, the framework does not provide an externally imposed normative account of "optimal" behavior in terms of, for example, reward maximization. While Friston's framework prescribes that the system minimizes free energy; *what* this minimization entails fully depends on the configuration of the anticipatory system. In other words, the specifics of free energy minimization are fully dependent on the contracted internal constitution of the anticipatory organism, and the surroundings in which this habitual structure dwells. For that reason, the only normativity involved in free energy minimization is that it emphasizes 'the necessary tendency of living organisms to resist the second law of thermodynamics; i.e., to maintain an internal structure or dynamics in the face of constant change' (Allen & Friston, 2018: 2473). Friston's approach to neural functioning thus describes and formalizes how, once a state of relative equilibrium is acquired – to the point where a maintainable boundary is established – the system organizes and produces itself through a perpetual loop of sensation and action.

Plasticity? Dialectics? Contradiction?

The etymological proximity of "habit" and "state" indicates that both in the case of Hegel and Friston, what is at stake is the organization of an anticipatory structure. In both cases, the organization of this structure takes the form

of a continuous process. Given a boundary, there is a process of boundary-maintenance. And given the process of boundary-maintenance, the structure constituted by the boundary is maintained. As such, the only stability that pertains to this structure consists in the perpetual maintenance of a separation between inside and outside. In the active process of boundary-maintenance, an organism under free energy minimization is no longer a passive processor of external influences, but does this mean that free energy minimization adheres to the notion of plasticity, as understood by Malabou?

Plasticity for Malabou designates the capacity to receive as well as the capacity to produce form. In the case of Hegel, we saw that the organism receives form from external nature, but it also produces form by positing external nature as part of its own organization. This process of self-production takes the form of a self-referential process through which the organism participates in its own determination. In the case of Friston's framework, we are dealing with exactly the same process. As we have seen, the organism minimizes free energy by actively limiting the states it occupies through a circular process implicating both sensation and action. As such, free energy minimization constitutes a loop in which the reception of form (sensation) and the production of form (action) are implicated in the same circular process. However, while the self-determination of sensation is a necessary requirement for the self-production of form, in order to speak of plasticity in a Hegelian sense we need a process which captures the reverberation of self-differentiation able to alter the contracted constitution of the organism's anticipatory structure. Does free energy minimization capture this process? Clearly it does; as the organism engages in perceptual and active inference, the anticipatory model itself is continually altered in confrontation with the disruption of the anticipations generated by this model. Under free energy minimization, the reception and production of form are implicated in the same process: 'learning and perception are two sides of the same coin': 'perceptual inference (i.e., neurodynamics) and learning (i.e., neuroplasticity) are in the game of optimizing the same thing; namely, model evidence or its variational equivalent (i.e., free energy)' (Friston, 2018b: 1020-1021). The only difference is that Friston's distinction between neurodynamics and neuroplasticity, fall under the Hegelian notion of plasticity. As such, free energy minimization is a plastic process. In this sense, Friston's framework moves beyond the one-sided conception of plasticity criticized by Malabou, but is this process dialectical?

What makes the process of plasticity dialectical for Malabou is that in it, contradictory operations coincide, in the sense that 'the seizure of form and the annihilation of all form, emergence and explosion, are contradictory' (Malabou,

2005: 12). The reception and production of form are not only subsequent stages in the formation of the organism, but both capacities actualize simultaneously. Every reception of form implies the simultaneous production of form as the organism incorporates the influences it undergoes. Put differently, whenever the organism converges on a state, the organism's overall state by definition encloses anticipation and the disruption of anticipation in the form of anticipatory error. In this sense, as in the case of Hegel, Friston's anticipatory organism "has itself for its object" and maintains a break not only with its surroundings, but with itself as well. It is only in the coincidence of the anticipatory organism's structure (identity) and its disruption ("error") that this structure is altered in anticipation of what to expect next: 'the threat of the explosion of form structurally inhabits every form. All current identity maintains itself only at the cost of a struggle against its autodestruction: it is in this sense that identity is dialectical in nature.' (Malabou, 2008: 71). In other words, the coincidence of (perceptual) inference and learning stems from the coincidence of anticipation with its negation. As such, the organism is simultaneously and necessarily subject to deformation in the process of its own formation. The convergence of Hegel and Friston shows how Friston's free energy minimization is a modern-day instantiation of Hegel's dialectical process of plasticity. Indeed, how 'it is Hegel who will have discovered before its discovery the plastic materiality of being: that free energy, whether organic or synthetic, which circulates throughout in each and every life' (Malabou, 2005: 193).

And yet, it is not enough to designate the formation-deformation of the organism's structure-process. Hegel's *Philosophy of Nature* does not stop there. We need to ground the unrelenting continuation of this process. Said differently, we need to ground the anticipatory organism's internal purposiveness. How? We saw that Hegel's designation of the organism ended with the primacy of contradiction constitutive of the organism's life: in passing through the tension between abstract and relational, we saw that the organism itself sustains contradiction between habitual structure and deficiency. We need to show how the movement of plasticity is sustained by 'the contradictory tension between particular determinacy and its dissolvement into the universal.' (Malabou, 2005: 12). It is here where Friston, and the appropriations of his framework in terms of cognitivism and enactivism, do not go far enough.

4.5. From tension to contradiction

The first question we set out to answer was the following: is free energy minimization a formalization of a dialectical process of plasticity, as understood by Malabou as the capacity both to receive and to produce form? Above we saw that this is indeed the case: both Malabou and Friston capture a self-productive process which simultaneously implicates *formation* as well as *deformation*. That brings us to the second question we posed: can we enact the Hegelian shift from tension to contradiction with regards to free energy minimization? What tension are we referring to?

There are two tensions that pertain to free energy minimization. The first is theoretical: the tension between cognitivism and enactivism; seclusion and openness. Cognitivism and enactivism are two theoretical frameworks that fundamentally differ in their approach to the mind and brain. Cognitivism dates back to the cognitive revolution of the 1950s, in which the emphasis shifted from behaviorism to the way the mind processes information (e.g., see Miller, 2003). Enactivism is a more recent approach which builds on ecological psychology, embodied cognition, and situated cognition. It is often contrasted with cognitivism, because enactivism places great emphasis on the dynamic interaction between brain, body, and environment, beyond mere information-processing (e.g., see Thompson & Varela, 2001). As we will see below, the appearance of Friston's framework differs substantially, depending on the perspective we adopt.

The second tension pertains to the organism itself: under free energy minimization, the anticipatory organism is subject to the unsolvable imperative to minimize free energy, in that it continually sustains anticipatory error. Our wager is that these tensions overlap: the tension between seclusion and openness *is the unsolvable tension of free energy minimization.* We will begin with the theoretical tension (Section 4.1), after which we will move on to the unsolvable imperative of Friston's framework (Section 4.2). Finally, we will attempt to enact the shift from tension to contradiction with regards to free energy minimization (Section 4.3).

Seclusion and openness

The theoretical struggle surrounding the appropriation of Friston's framework revolves around 'the tension between internalist and externalist approaches' (Allen & Friston, 2018: 2478). This tension revolves around the following question: is the organism separated or secluded from the surroundings in which

it dwells, or should the emphasis lie with the organism's embeddedness in and perpetual openness to these surroundings?

The first position is defended most forcefully by Jacob Hohwy:

> 'All perceptual and active inference happens in an interplay between the evidence to the system, that is, activity at the sensory epithelia, and the predictions generated under the overall model in the brain. This creates a sensory blanket—the evidentiary boundary—that is permeable only in the sense that inferences can be made about the causes of sensory input hidden beyond the boundary.' (Hohwy, 2016: 265)

From this perspective, the emphasis lies fully with the seclusion of the internal anticipatory model within the confines of the skull. All of the organism's activity (bodily or otherwise) stands in service of the optimization of this model. As such, free energy minimization becomes 'more neurocentrically skull-bound than embodied or extended, and action itself is more an inferential process on sensory input than an enactive coupling with the body and environment' (Hohwy, 2016: 259).

The second position – emphasizing the brain and body's embeddedness/perpetual openness – is formulated succinctly by Jelle Bruineberg and Erik Rietveld:

> 'The FEP [free energy principle] implies a deep connection between the dynamics of the brain-body-environment system and the neurodynamics. [...] The function of the generative model is therefore not to provide the agent with a representation of the dynamical structure of the environment per se, but rather to steer its interactions with its environment in such a way that a robust brain-body-environment system is maintained. (Bruineberg & Rietveld, 2014: 7)

From the enactivist position, the skull-secluded internal model stands in service of the organism's selective openness towards the affordances provided by the environment. Due to this enduring relation between the organism and its surroundings, the 'organism does not need to have a model of its niche, but rather the claim is that the structure of the niche is reflected in the structure of the skilled embodied organism' (Bruineberg & Rietveld, 2014: 8). In other words,

the organism 'does not have a model of its world – it *is* a model.' (Friston, 2013a). How to resolve the tension between cognitivism and enactivism?

Not only do both positions have merit, but they follow logically from free energy minimization, and neither are naïve. The enactivists recognize the requirement for a distinctive entity in order for there to be a 'brain-body-environment system': '[o]ne might object that there is still a non-trivial boundary separating the system from its environment [...]. We agree [...].' (Bruineberg et al., 2016: 2438). At the same time, Hohwy is well aware that the skull-secluded internal model is simultaneously open to its environment. For him, the 'challenge is then to balance seclusion and openness in our understanding of the mind-world relation' (Hohwy, 2016: 266).

For Friston himself, 'the FEP [free energy principle] resolves the tension between internalist [cognitivist] and externalist [enactivist] approaches' (Allen & Friston, 2016: 2478). The solution takes the form of "a little bit of both":

'Clearly, the active inference account satisfies the criteria for a radically embodied theory of mind. According to the free energy principle, an organism is best understood as a system of mutually interlocking systems; the body, mind and environment are inextricably bound up in the organism's free energy minimization: in fact, all the heavy lifting done by active inference is in preserving a degree of (statistical) separation between the body, mind and environment' (Allen & Friston, 2018: 2475-2476)

Although ultimately, 'the FEP offers a formal path forward for enactivism. By providing a guideline to discovery, the normative principles embedded within the approach allow enactivists to go beyond arguing about the demarcations of the organism.' (Allen & Friston, 2018: 2478). In other words, what we get is a balanced solution with a little bit more enactivism than cognitivism. The question is whether "finding the right balance" is the best we can do; whether the tension between cognitivism and enactivism is to be resolved at all.

Tension redoubled

What makes free energy minimization such a compelling framework is that it is grounded in the necessary preconditions for the existence of the organism (Allen & Friston, 2018: 2473). If this claim is justified, and if the cognitivist and enactivist accounts follow logically from it, should we not consider whether the tension between both positions pertains to the anticipatory organism itself? Despite the 'deep reciprocity between the embodied and environmental facts of the organism' (Allen & Friston, 2018: 2475), and the endless variations on transient extended cognitive systems (Clark, 2016a), 'the very existence of a system mandates the separation between the system and its external

milieu' (Allen & Friston, 2018: 2473). As such, the tension between cognitivism and enactivism is not only maintained, but we can only get the extended, enactivist account through the cognitivist emphasis; through the imposition of a boundary. It only makes sense to speak of the organism on account of the existence of a minimal separation between organism and surroundings. Is this not the same problem we encountered in Hegel? Are we not again dealing with the tension between an abstract and a relational conception; a contradiction between independence and dependence?

We can easily reframe the tension between cognitivism and enactivism in terms of the stages Hegel distinguished. The cognitivist conception of the minimization process corresponds to the organism understood as a 'system of sensibility', which subsequently is differentiated outwards through its sensory and motor nerves. These are the moments that Hegel calls sensibility and irritability, respectively. From this perspective, the organism appears as a self-subsistent entity, where all the emphasis is placed on the internal organization of the organism undergoing change in a perpetual process of reproduction. Against this abstract designation, enactivists emphasize the necessity of the relation that the organism maintains with its outside, through which the organism engages in mutual exchange with its surroundings. And yet, the necessity of this relation depends on the existence of a self-subsistent entity. Without seclusion, there would be no open exchange, since there would be no system to engage in exchange. The paradox is that the possibility of perpetual exchange only arises from a situation in which an organism maintains a boundary between itself and its surroundings. The other way around also holds: the organism maintains its seclusion only by engaging in open exchange with its outside. Thus, by maintaining their precarious minimal degree of autonomy, it implies that 'organisms, by being organizationally closed, are also necessarily thermodynamically open' (Marques & Brito, 2014: 99). However, as in the case of Hegel, we are not just dealing with tension between conceptual formulations: Friston's anticipatory organism itself sustains tension.

For the organism, there is no way to establish the accuracy of its anticipations. All the organism can do is minimize the discrepancy between its anticipations and its elicited sensory states, by either changing its anticipatory model, or its sensory states through action. In Friston's mathematical formalizations, the minimization problem revolves around the difference between probability distributions already present on the one hand, and inferred probability distributions based on new input on the other. Indeed, 'this difference is always positive' (Friston, 2010: 128). Said differently, since minimizing free energy 'places an upper bound on the entropy or dispersion

of sensory states' (Friston, 2012: 2), the implication is that the system is continually subject to dispersion in need of bounding. The organism cannot let its sensory states disperse indefinitely if it is to maintain a minimum of consistency. At the same time, because the organism operates from within its own bounded organization, without direct access to its hidden external states, the organism also cannot get rid of dispersion entirely. Concretely, dispersion is nothing but the anticipatory error the organism sustains in the convergence on every state it takes up. As the organism moves through different states, each configuration is simultaneously marked by error; each anticipation is subject to disruption. The continuous state of tension under free energy minimization resides in the inseparable combination of anticipatory state "plus" disruption. Two questions need to be asked at this point. First, where is this tension located? Are we dealing with tension between the organism and its surroundings, or is it a tension within the organism itself? Second, how to conceive of this tension? Is its solution a goal the organism strives towards, or is it a problem concurrent with and constitutive of the organism's existence?

Contradiction again

While both cognitivists and enactivists recognize the necessity of an enduring tension under free energy minimization, they differ in their conception of this tension. For enactivists, the organism maintains an 'ever present dis-attunement between environmental dynamics and internal dynamics' (Bruineberg et al., 2016: 5). At first glance, this formulation makes perfect sense. After all, the organism needs to maintain itself in the face of external nature. We found a similar formulation in Hegel, where the organism is directed towards the outer world 'as being inwardly in a state of tension towards it' (Hegel, 2004: 381, §357). The crucial difference is that in the case of Hegel, the organism is *inwardly* in a state of tension towards its surroundings. While it is of course true that the organism stands in tension or dis-attunement toward its surroundings, external nature does not act directly on the organism's internal organization, because this internality is shielded by a boundary comprised of sensory and action states. Since the relationship between the organism's internal organization and its external states is mediated by this boundary, there is no tension operative directly between the organism and its surroundings; between internal and external dynamics. Instead, this tension is sustained internal to the organism's organization; namely, between the organism's internal anticipatory model and its sensory states. As hairsplitting as this distinction may seem, it is crucial, as we will see below.

From the perspective of cognitivism, both the unsolvable status, as well as the internality of the minimization problem is recognized: the 'prediction error or free energy bound on surprise is never zero.' (Hohwy, 2013: 172). As such, the organism 'is engaged in an internal struggle to make its states fit with its input' (Hohwy, 2013: 179). However, from this perspective, the minimization problem becomes a 'moving, ultimately unobtainable goal' (Hohwy, 2013: 174). The question is whether this formulation is justified. Can we say that the organism strives towards resolving its own minimization problem? What would such a solution entail? Can there be a living organism that is not subject to the imperative to minimize free energy?

With these questions we touch on the paradoxical status of free energy minimization. There is no room in Friston's framework for a grand conclusion where the organism succeeds definitively in the minimization of free energy. At least, not if the organism is to maintain itself as a distinct entity. From the perspective of Friston's framework, a bounded organization necessarily engages in free energy minimization, but for minimization to take place there needs to be a bounded organization. For the living organism, this means that the only way to escape from the imperative to minimize; to escape from this state of tension, is to give up its bounded organization. In other words, death is the only definitive solution to the problem of free energy minimization. The paradox is that an unresolvable state of tension arises concurrently with the imposition of a boundary; with the precondition for the organism's existence. It is because of the concomitance of boundary and tension that we need to invert our perspective, to the point where we repeat the procedure of Hegel's philosophical exposition with regards to Friston's framework; in order to enact the shift from unsolvable tension to contradiction constitutive of the organism's life.

With this shift we return to the opposition between external and internal purposiveness, and the problem of natural ends. In the same way that the tension between cognitivism and enactivism is not a sole theoretical tension, but simultaneously operative at the level of the organism itself; the opposition between external and internal purposiveness does not pertain only to the level of explanation. We are not just dealing with the problem of how to ground an explanation of the organism in the preconditions for the existence of that organism; the problem is equally that which grounds the organism's perpetual anticipatory activity. In the same way that the organism in Hegel sustains contradiction between habitual structure and deficiency, Friston's anticipatory organism sustains contradiction between internal states and dispersion. In both the case of Hegel and Friston, the role of dispersion and deficiency is that

of lack positivized: an interruption which nonetheless plays a "positive" role in the internally directed revision, or externally directed activity it drives.

To take Hegel's example of hunger, it is on account of the organism's internal deficiency that it seeks out, ingests, and digests food. If the organism is to maintain its life, it cannot let its deficiency run amok; if it does not address its hunger, it will die. In this sense, the Hegelian organism also bounds deficiency. The great advantage of free energy minimization is that the basic logic of Hegel's elementary example of hunger is truly 'particularized in an infinite variety of ways' (Hegel, 2004: 388, §360). With Friston's framework, the basic logic of Hegel's dialectical process is generalized to every aspect of the organism's perceptual and motor capacities, whereby the entirety of the organism's internal structure "feeds off" its surroundings. If Hegel's contradictory organism is a process that revolves around its own deficiency, then free energy minimization is a process which subsumes and bounds all of the organism's deficiencies. Not only do both free energy minimization and activity of deficiency pertain to a perpetual process of self-organization and self-production, but free energy minimization *is* activity of deficiency.

Due to the recurrence of deficiency, and the concomitance of this recurring problem with the existence of a boundary, it is insufficient to formulate the minimization of free energy in terms of an unobtainable goal. For the organism, there is no unobtainable goal, because the "goal" consists in sustaining the continuous process itself, through which dispersion or lack inherent to the organism's organization is constrained, and a boundary is maintained. We could say the "goal" is continually "reached" and "missed" simultaneously. A definitive solution to the problem of free energy minimization is repeatedly missed, but this miss is the goal itself. More precisely, we are dealing with the 'splitting between goal and aim, the moment when the true aim is no longer to hit the goal but to maintain the very circular movement of repeatedly missing it.' (Žižek, 1993: 199). In this sense the organism exists as a natural end: the process of free energy minimization is without end, because the end is the process of free energy minimization itself. As such, the direction of boundary-maintenance is not outwards: in the attempt of biological systems to maintain their states and form, we are dealing with a process that is thoroughly self-directed.

In addition to the designation of the organism's self-movement in terms of self-organization, self-production, and self-determination, perhaps it is self-limitation which best captures the convergence between Hegel and Friston, and simultaneously announces their minimal, but crucial difference. As the organism revisits a limited number of states that lie within its physiological

bounds, it engages in a limiting movement of the "self" directed at the "self": 'life emerges when external limitation (of an entity by its environs) turns into self-limitation.' (Žižek, 2009: 205). This is the process that is central both in the case of Hegel and Friston: everything hinges on the existence and maintenance of a boundary. No boundary, no organism.

At the same time, self-limitation is operative in terms of a limitation (lack) at the level of the "self":

'Where there is a limitation [*Schranke*], it is a negation only for a third, for an external comparison. But it is *lack* only in so far as the lack's overcoming is equally present in the same thing, and contradiction is, as such, immanent and explicitly present in that thing.' (Hegel, 2004: 385, §359)

While both Hegel and Friston emphasize the necessity of an organization that is bounded [*beschränkt*], only in Hegel we find the primacy of contradiction constitutive of the process that sustains such bounded organization. If we enact the shift from tension to contradiction, the unsolvable status of the minimization problem functions not as an unobtainable goal, but serves first and foremost as the condition both for the organism's existence as well as its perpetual free energy minimization. Only in trying to overcome the unsolvable obstacle of its persistent deficiency or dispersion does the organism continue to anticipate: it is this continually present inherent limitation around which the organism's anticipatory activity circulates. For this reason, the tension between organism and surroundings needs to be transposed back into the organism, to the point where we conceive of this tension as the organism's constitutive contradiction.

To return to the opening problem of this section: the tension between internalist and externalist approaches is not only retained; it is elevated to a contradiction constitutive of the perpetual process of free energy minimization. If the organism's founding gesture is the imposition of a boundary, then this gesture simultaneously condemns the living being to a life of sustained contradiction. This is the fundamental dialectic of Friston's free energy minimization expressed in the opposing positions of cognitivism and enactivism.

This perspective dissolves the so-called "dark-room problem": the question concerning why, if an organism is primarily concerned with the minimization of prediction error/free energy, it does not simply seek out a dark and silent place: if there is no input to the system, then there is nothing to minimize. This is a problem only if we conceive of the organism as a fully autonomous and constituted entity, which subsequently engages in minimization. From a Hegelian perspective, it is a problem only if we remain at the level of an

abstract designation in terms of sensibility and irritability. If instead, we follow Hegel to the very end, the organism's constitutive contradiction appears equally necessary for the organism's existence; no less so than the existence of a boundary. From this perspective, the organism's activity is grounded not in some positive aspect of its organization, but in the sustained negation of this organization, around which the organism's anticipatory self-production circulates (see Fig. 1). In dealing with the dark-room problem, it is not enough to emphasize the need for continued scientific work; how, in 'due course, realistic working models will be forthcoming, at which stage this philosophical debate will rightly give way to detailed empirical evaluation of the claims being made.' (Friston et al., 2012: 6). While Friston's declaration and willingness to abandon his framework in the face of failure is admirable (a stance admittedly missing in Hegel, see (Žižek, 2012: 462); perhaps there won't be a need to 'search for a better model!' (Friston et al., 2012: 6). Hegel's philosophy is well able to dispose of "philosophical debates" such as the dark-room problem on its own. And if we enact the Hegelian shift from tension to contradiction with regards to free energy minimization, so is Friston's framework.

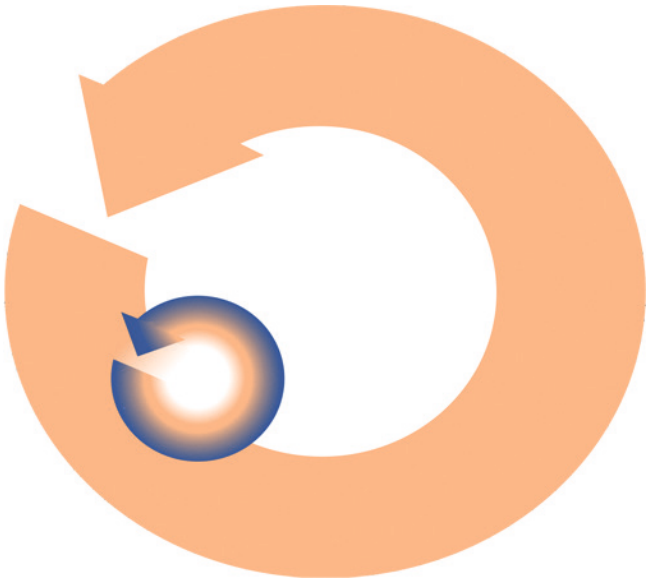


Figure 1. Schematic presentation of the contradictory organism. Every organism (blue) operates within the dynamics of its surroundings (brown). The organism's structure as distinct from its surroundings is maintained only as a continuous process of self-organization and self-production, revolving around recurring lack or dispersion (white) internal to the organism's organization (Cf. Friston, 2005, Figure 1). The organism's internal habitual model is a representation of, and "feeds off", its surroundings.

4.6. Discussion

This paper sprang from the convergence between Hegelian dialectics and Friston's approach to the brain. We set out to answer two questions. First, is free energy minimization a formalization of the dialectical process of plasticity, as understood by Malabou? We saw that indeed, in the coincidence of free energy minimization and the Hegelian notion of plasticity, Friston provides us with an approach to the brain which reinvigorates Hegelian dialectics from the perspective of neuroscience. In both cases, we are dealing with a perpetual process in which form is received and produced, and where *formation* and *deformation* occur simultaneously. Whether we speak in terms of a habitual anticipatory structure, or a model comprised of anticipatory states, what is important is that in both cases, the organism does not *have* a model, but *is* a model; there is no organism outside the process wherein its anticipatory structure is continually reproduced and transformed.

The second question central to this paper was this: can we enact the Hegelian shift from tension to contradiction with regards to free energy minimization? This question touched on two tensions simultaneously: theoretical and actual. In Hegel's work, we saw that the determination of the organism took the form of a succession of stages, passing from abstract to relational; from independence to dependence. However, the tension between these designations was not resolved. On the contrary, this tension needed to be located and mobilized in the actual workings of the organism itself. This is where we encountered the shift from tension to contradiction. If tension pertains to an opposition *between* concepts (e.g., seclusion and openness) or entities (e.g., organism and surroundings), then contradiction is an antagonism internal to one of the terms. In our case, we saw that this contradiction pertains to the organism's simultaneous maintenance of identity, as well as the negation of this identity in the form of recurring lack or deficiency. The attempt to answer the second question central to this paper could be viewed as nothing more than the attempt to repeat Hegel's procedure with regards to Friston's free energy minimization. First, we regarded the conceptual tension between cognitivism and enactivism; between the emphasis on seclusion and openness. We tried to cast this tension as the succession from an abstract to a relational conception of the organism. Second, instead of trying to resolve this tension, we shifted our attention to the organism's actual tension under free energy minimization; namely, the organism's unrelenting imperative to minimize entropy, dispersion, or anticipatory "error". In this sense, both "lack" in Hegel, and "dispersion" in Friston, serve the same structural role: they threaten the

perpetuation of the organism's organization and as such, must be kept within bounds. Finally, because the organism's contradiction between identity and negation exists concurrently with the existence of a boundary, it is constitutive for the organism's existence and at the basis of the organism's perpetual anticipatory activity. As such, the shift from tension to contradiction allows us to see how Friston's free energy minimization is a modern instantiation of Hegelian dialectics.

While we appealed to Hegel in order to bring to light the contradictory tension at the centre of free energy minimization, we should emphasize that Friston's framework goes beyond Hegel in the generalization of the rudimentary example of hunger to the organism's functioning in the broadest possible sense, to the point where the logic of plasticity pervades every aspect of the organism's functioning. Furthermore, Friston's framework goes beyond Hegel in the specification of the concurrent processes of anticipation and precision estimation. Not only is the organism an anticipatory structure, but the transience of this structure depends on the continuous interplay between the anticipatory state, the precision of this state, as well as the precision afforded to the disruption of this state. As such, free energy minimization formalizes the mechanism that determines an organization's plasticity, in which the Hegelian legacy is maintained.

Chapter 5

Learning to predict based on self- versus externally induced prediction violations: a direct comparison using a Bayesian inference modelling approach

---

In preparation as:

Boonstra, E. A. & Slagter, H. A. (in prep.). Learning to predict based on self- versus externally induced prediction violations: a direct comparison using a Bayesian inference modelling approach.

5.1. Abstract

Predictive processing is quickly gaining ground as a theory of perception and attention. From this perspective the brain is cast as an organism's predictive model of how its world works and will continue to work in the future. However, research on the brain's predictive capacities remains beholden to traditional research practices in which participants are passively shown stimuli without their active involvement. The current study is an investigation into ways in which self-generated predictions may differ from externally induced predictions. Participants completed a volatile spatial attention task under both conditions on different days. We used the Hierarchical Gaussian Filter, an approximate Bayesian inference model, to determine subject-specific parameters of belief-updating and inferred volatility. We found preliminary evidence in support of self-generated predictions incurring a larger reaction time cost when violated compared to predictions induced by sensory cue, which translated to participants' increased sensitivity to changes in environmental volatility. Our results suggest that internally generated predictions may be afforded more weight, but these results are complicated by session order and duration effects, as well as a lack of statistical power. We discuss the limitations of our study preventing us from replicating previous research, and ways to remedy these shortcomings in future studies.

5.2. Introduction

The study of perception and attention in living organisms is most often conducted from an outside-in perspective, from which primacy is afforded to information and stimuli impinging on sensory epithelia. For example, tasks are used in which external cues tell participants where to attend. This perspective or approach becomes problematic if we recognize that living organisms maintain internal dynamics of their own, based on which they perceive, attend, act, and operate. Indeed, it has been argued that an inside-out perspective, where an organism's internal states are afforded primacy, is a more appropriate starting point for the study of brain and behavior (Buzsáki, 2019).

An inside-out perspective is inherent to research on self-elicited action-effects. Already in the 1970s, the seminal study by Schafer and Marcus (1973) reported that self-delivered auditory and visual stimuli attenuated the amplitude of event-related potentials (ERPs) by respectively 50% and 28%, compared to when stimuli were machine-delivered. Since then, research on the directional relationship from action to perception has been framed in terms of agency; revolving around the question how and to what degree the system ascribes the cause of sensation to itself, in contrast to unsolicited stimulation from outside (Synofzik et al., 2013; Wen, 2019) and action-effects (Elsner & Hommel, 2001; Hommel, 1996, 2019; Hommel & Elsner, 2009; Prinz, 1997). In the auditory domain in particular, the finding that self-generated tones are accompanied by attenuated early ERPs such as the N1 and P2 has been replicated many times (Korka et al., 2022), but the behavioral effect is less conclusive: perceptual sensitivity has been found to both decrease (Weiss et al., 2011; Weiss & Schütz-Bosbach, 2012) and increase (Reznik et al., 2014; Reznik & Mukamel, 2019). While the auditory domain is more extensively studied, similar neural effects have been observed for the visual domain (Cardoso-Leite et al., 2010; Nittono, 2004; Nittono et al., 2003; Vasser et al., 2019; but see Schwarz et al., 2018). While the exact workings of the relationship between action and perception are unclear as of yet, it is clear that stimuli stemming from self-generated predictions are processed differently from externally induced ones.

An inside-out perspective is also part and parcel of predictive processing approaches to the study of brain and behavior (Clark, 2013; Friston, 2010; Hohwy, 2013). From these perspectives, the brain instantiates an organism's predictive model of its outside world, which upholds expectations regarding how this world works and how it will continue to work in the future. This model undergoes changes as the organism is confronted with violations to its expectations; violations it tries to minimize overall. This process of model-

updating and prediction error minimization can be cast as approximate Bayesian inference (Da Costa et al., 2021; Knill & Pouget, 2004; Pouget et al., 2013). From this perspective as well, the violation of internally and externally generated predictions are thought to be weighted differently (H. Brown et al., 2013), but to our knowledge no study has made this comparison directly for the visual domain within the context of Bayesian predictive perceptual processing.

We set out to test this difference within the context of the Hierarchical Gaussian Filter (HGF); a computational model implementing approximate Bayesian inference linking three hierarchically ordered levels capturing not only trial-by-trial probability estimates, but also the inferred volatility of an agent's environment (Mathys et al., 2011; Mathys et al., 2014). We rely heavily on earlier work in which this model was used to show how model-updating changed over time in the context of a spatial attention task (Vossel, Bauer, et al., 2014; Vossel, Mathys, et al., 2014). In these studies, a cue indicated whether a target would appear left or right of fixation. Crucially, the probability with which the cue was valid changed over the course of the experiment across three levels (50, 69, 88%) unbeknownst to participants. It was found that participants were not only faster with eye movement responses on valid compared to invalid (e.g., cue indicating left, target appearing right) trials, but this effect was stronger on blocks where the probability of a valid cue was larger. In terms of modelling, they found an optimal model aligned with the proposal by Feldman and Friston (2010), suggesting that in general the precision of prediction errors drive changes in model-updating, to the point that higher precision results in more rapid updating, and where in the case of this particular model, the trial-by-trial precision estimate of the first-level prediction drives responses. The importance of contrasting this setup with an action condition in which predictions are self-generated comes from the idea that through action, a predictive system can actively test the validity of its internal model. Instead of being at the whim of external stimulation (passively awaiting what comes in), the system can generate sensory information to actively test its models, facilitating the speed at which reliable models can be produced.

Our main research question is as follows: how does the response to induced predictions and their violation differ when these predictions stem from externally generated sensory cues compared to internally generated action?

We first report a pilot study in which we transposed the experimental design by Vossel and colleagues (2014) to an action-effect paradigm. Our pilot was the same as the original study, except that spatial predictions were self-generated by participants' own actions instead of external cues. We expected

to replicate Vossel and colleagues (2014) by finding slower eye movement responses for invalid compared to valid trials, and that this difference would be larger for higher probability blocks. Like Vossel and colleagues (2014), we expected an optimal "precision" response model to account for these findings. An indirect comparison with Vossel et al.'s findings showed that the RT cost for violated predictions were larger when this prediction stemmed from self-generated action, in our pilot study, which shaped our hypotheses for our main within-subject experiment in which participants completed an action and cue version of the same task on different days. This allowed us to critically, directly test if these conditions differ by keeping constant as many confounds as possible.

We expected to replicate and extend our pilot finding; that the violation of predictions through invalid trials would incur a larger reaction time cost when a prediction stems from action rather than from a sensory cue. Moreover, in terms of HGF modelling, we expected to see these effects reflected in a larger belief-updating parameter (ω_2, see below) for the action condition. This would support the notion that perceptual predictions are afforded more weight when they are self-generated.

5.3. Methods

Participants

Twenty participants were recruited from the Vrije Universiteit Amsterdam subject pool for the pilot, while thirty-five participants were recruited for the main experiment. After exclusion based on pre-specified criteria (see below), the final sample for the pilot consisted of 19 participants (mean age 21.2, range 18-31 years, 16 women), while for the main experiment 25 remained (mean age 19.8, range 18-23 years, 20 women). All participants reported normal or corrected-to-normal vision, provided written informed consent prior to participation, and received either a monetary reward or course credit for participating. The experiment was approved by the ethics review board of the Faculty of Behavioral and Movement Sciences of the Vrije Universiteit Amsterdam.

Procedure and experimental paradigm

In the pilot study, participants performed a spatial predictability task, in which the predictability of the sensory outcomes of their own actions (a button press) varied across trials. In the main experiment, participants came to the lab for two sessions, and either performed this task or a task in which the spatial predictability of a visual cue was varied across trials (cf. Vossel et al., 2014).

In both tasks, participants had to move their eyes as quickly and as accurately as possible to a target appearing either to the left or right of a central fixation cross. The design and analyses of the main experiment were preregistered (https://doi.org/10.17605/OSF.IO/V6DJ2), unless otherwise noted.

Stimuli were presented on a CRT monitor (1680 ×1050 pixels, 75 Hz). Eye movements were recorded using an EyeLink 1000 Plus eye tracker (SR research, Ontario, Canada). Participants viewed the screen through a chinrest at a distance of 70 cm.

Participants were instructed to fixate centrally on an ABC fixation dot as defined by Thaler and colleagues (2013) (.55° in size) until a target appeared. Once the eye tracker registered fixation, the trial would start 500 ms later by the presentation of two squares on the horizontal midline (1.9° in size, 8.1° from fixation) to the left and right of the fixation dot, indicating to participants either an impending cue or that they were to initiate an action, depending on the condition. In the action condition, participants initiated trials by freely choosing one of two buttons ("s" and "k" on a standard keyboard) with their index fingers. The left ("s") button most often gave rise to a target on the left, while the right ("k") button had the same function for a target on the right, exploiting a natural spatial association (Leuthold, 2011; Simon & Rudell, 1967), also present for the left and right arrow head cues used in the cue condition (Figure 1a). Participants were instructed to use both buttons approximately evenly, and to avoid simply alternating between buttons. On average participants took 534 ms (SD: 134 ms) to initiate an action. In the cue condition, a cue was presented after 500 ms for 200 ms without the participant's involvement. 800 ms after an action or a cue, a target in the form of a gabor patch (1.3° in size) appeared in one of the two placeholders to which participants were instructed to move their gaze as quickly and as accurately as possible. Once participant's gaze reached the target, or after 1700 ms, the trial would end and the next trial would start (see Figure 1).

One session took approximately 1.5 h, in which participants completed a practice block of 36 trials, as well as 612 experimental trials, evenly divided into 18 blocks of either 32 or 36 trials. Trials in the cue condition were interspersed with 108 "null-trials", where only the baseline display (fixation and squares) were shown to interrupt the flow of trial presentation (cf. Vossel et al., 2014). Crucially, blocks of trials varied unbeknownst to participants in how likely the target was to follow either cue or action. Said differently, when a cue indicated left or when a participant pressed the left button, it was uncertain whether the target would appear on the left side. This validity probability varied block wise

at 88, 69, and 50%. The practice block was steady in terms of validity probability at 88%. We used a within-subject design where all participants completed two sessions on different days with at least one day in between. The pilot consisted only of the action session.

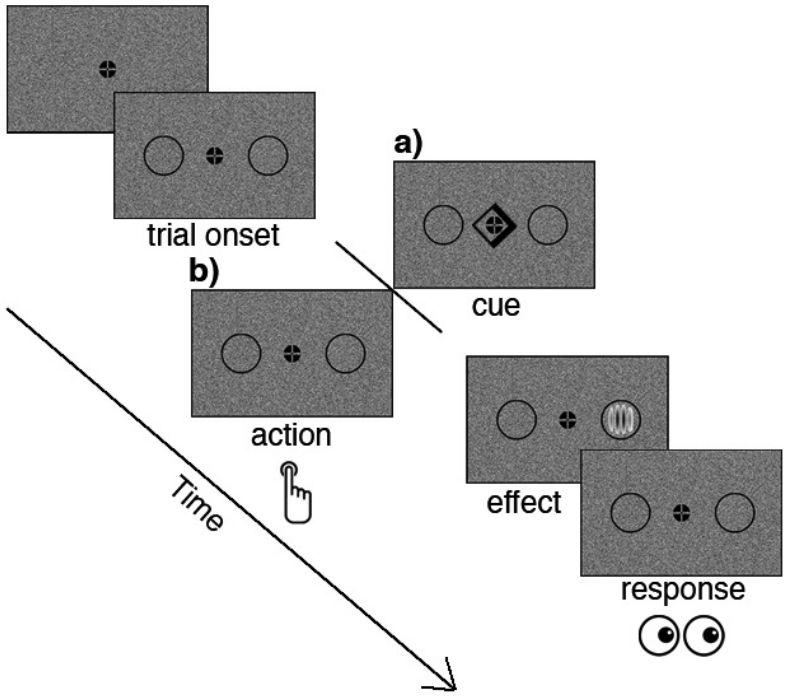


Figure 1. Experimental procedure for one trial in the **a)** cue and **b)** action condition. Depending on the condition, a cue or an action predicted whether a target was likely to appear left or right of fixation or equally often at either location, and participants had to move their gaze to the target when it subsequently appeared.

Eye tracking and preprocessing

We used the same analysis pipeline for all the data reported in this paper. All data were analyzed offline. In line with Vossel and colleagues (2014), the start and endpoints of saccades were defined using a velocity-based algorithm, except we used the data-driven method developed by Nyström and Holmqvist (2010) to account for individual differences and thereby maximize sensitivity for saccade detection. We calculated saccade latency (time between target presentation and the start of the first eye movement) and landing position of the first saccade for every trial. In line with Vossel and colleagues (2014), only the first saccade was analyzed. A target was considered reached when participant's gaze traversed at

least two-thirds of the distance towards the target for at least 10 ms. Trials were excluded when a saccade started out more than 1° from fixation, when saccadic RT was smaller than 90 ms, when less than two-thirds of the distance was traversed to one of either target locations, when no saccade could be detected, when a trial contained more than 20% missing values, and when a response was incorrect. Individual participants were excluded from analyses entirely when we found more than 50% of their data to be missing. In this way we excluded one participant for the pilot and two for the main experiment to arrive at our final samples.

Data analysis

Pilot experiment

Our main dependent measure was response speed (RS): the inverse of saccadic reaction time due to its normal distribution (Vossel, Mathys, et al., 2014). To verify that the probability and validity manipulations worked also when predictions were internally generated through action, we ran a 3 (Probability; 88/69/50%) x 2 (Validity; valid/invalid) repeated-measures ANOVA with RS as the dependent measure.

Main experiment

To test our main question, whether internally generated predictions differ from externally induced ones, we conducted, as preregistered, a 3 (Probability; 88/69/50%) x 2 (Validity; valid/invalid) x 2 (Expectation; action-/cue-induced) repeated-measures ANOVA with RS as the dependent measure. We computed two additional ANOVAs to control for possible confounds. One with the between-subject variable Session Order added, and one in addition to our preregistration, but in line with Vossel and colleagues (2014), with the within-subject variable Time (first/second half of the experiment). We report Greenhouse-Geisser corrected p-values when the assumption of sphericity was violated, as indicated by decimal denoted degrees of freedom.

To evaluate evidence in favor of our (null) hypotheses, we conducted Bayesian statistics. For each reported frequentist ANOVA, we report the Bayes factor corresponding to the inclusion of a factor or interaction within the model in question (shortened to BF_{incl}), compared to equivalent models stripped of the effect. For example, $BF_{incl} = 10$ indicates that a model including the factor in question is ten times more likely given the data compared to a model without the variable. Conversely, $BF_{incl} = .1$ indicates that a model without said effect is ten times more likely given the data. In addition to frequentist t-tests, we report

the Bayes factor corresponding to the relative likelihood of a difference between conditions versus no difference (shortened to BF_{10}). All Bayesian statistics were conducted using JASP version 0.16.2 (Love et al., 2019).

Data visualization was performed with the help of raincloud plots (Allen et al., 2019), which include the mean, individual data points, as well as the overall distribution of the measure in question.

HGF

To formalize subject-specific trial-by-trial belief updating, we employed a Hierarchical Gaussian Filter (HGF) as implemented in the TAPAS toolbox (http://www.trans lationalneuromodeling.org/tapas/) (Frässle et al., 2021; Mathys et al., 2011; Mathys et al., 2014) for MATLAB (2019a, The MathWorks, Inc., Natick, Massachusetts, United States). The HGF is a Bayesian hierarchical learning model consisting of a perceptual part in which a probability is estimated (in the case of the current study the probability of a valid cue/action), and a response part in which this probability estimate is transformed into RS (see Figure 2). An important advantage of this model is that because it implements approximate (or variational) Bayesian inference, it avoids the computationally costly procedure of Markov chain Monte Carlo (MCMC) sampling, a commonly used Bayesian modelling framework (cf. Wagenmakers & Lee, 2014).

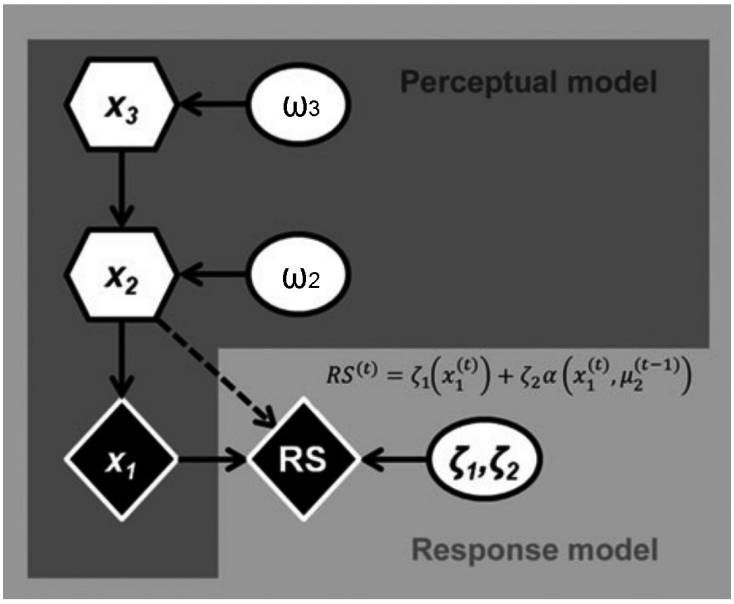


Figure 2. The hierarchical gaussian filter (HGF). Adapted from Vossel and colleagues (2014).

The first level ($x_1^{(t)}$) of the perceptual model comprises a Bernoulli distribution from which the trial-by-trial probability estimate of a valid cue/action is derived. This level is governed by the level above ($x_2^{(t)}$). The pace at which this second level changes is governed by the third and final level above ($x_3^{(t)}$) as well as a subject-specific fixed parameter ω_2. The step size at which the third level changes is governed by a second subject-specific parameter ω_3 (denoted as ϑ in earlier implementations of the HGF (Mathys et al., 2014; Vossel et al., 2014)). Both the second and third level change as a random Gaussian walk. Said differently, ω_2 reflects the speed of belief updating about trial-by-trial cue/action validity, while ω_3 reflects the speed at which the stability of action/cue validity is updated.

Subject-specific beliefs about trial-by-trial cue/action validity and volatility (posterior densities in relation to the hidden states $x^{(t)}$) are inferred from observable behavior (measured RT) by inverting the perceptual model. The sufficient statistics of the subject-specific posterior beliefs are denoted by $\mu^{(t)}$ (mean), $\sigma^{(t)}$ (variance), and $\pi^{(t)} = 1 / \sigma^{(t)}$ (precision).

The response model captures how an agent's belief translates to a decision (Daunizeau et al., 2010), or more specifically how the probability estimate at the

first level ($x_1^{(t)}$) translates to response speed. The response model comprises two parameters ζ_1 and ζ_2 which reflect the intercept and slope, respectively, of the linear relationship between the attentional factor α and posterior belief μ_1 (see Figure 2). This attentional factor reflects a proportion of attentional resources allocated to one of two stimulus locations. Three response models were defined by Vossel and colleagues (2014) which differed in how α was computed: 1) a "belief" model in which RS is a linear function of the probability estimate at $x_1^{(t)}$, 2) a "precision" model where α was determined by a sigmoid transformation (s) of $\pi^{(t)}$, the precision of the prediction at the first level (H. Feldman & Friston, 2010; Vossel, Mathys, et al., 2014), and 3) a "surprise" model where α is a nonlinear function of Shannon surprise (Bestmann et al., 2008; Vossel, Mathys, et al., 2014). See Figure 3 for the relationship between α and the probability estimate μ_1 at the first level. Out of these three models, we selected the most likely one given our data through Bayesian model selection (Stephan et al., 2009).

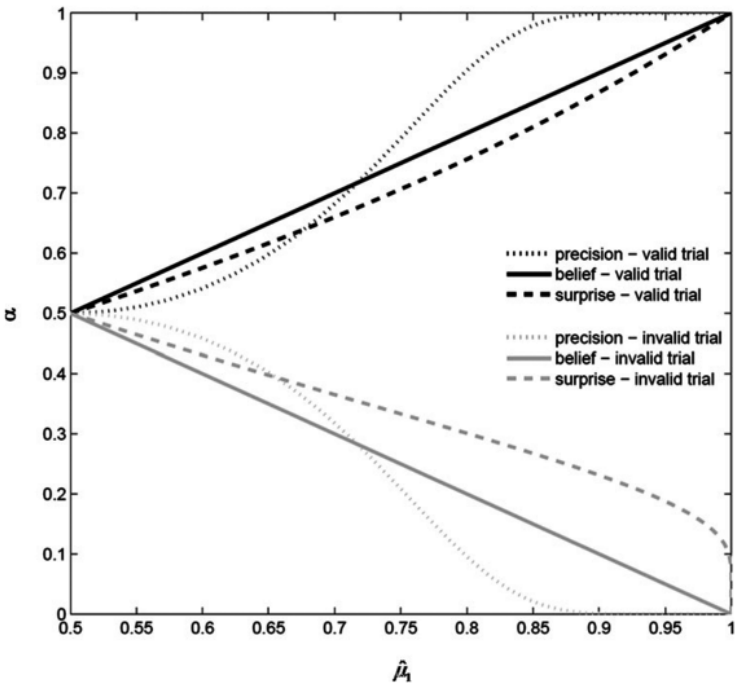


Figure 3. The relationship between attentional factor α and posterior belief μ_1 for all three response models. Adapted from Vossel and colleagues (2014).

5.4. Results

Pilot

For the pilot, 22.9% of data was excluded from further analyses. Similar to Vossel et al (2014), 19.3% of trials were excluded because the saccade started more than 1° from fixation, 1.1% because saccadic RT was smaller than 90 ms, 1.8% due to less than two-thirds of the distance being traversed to one of either target locations, .09% because no saccade was detected, .4% because trials consisted of more than 20% missing values, and .19% due to incorrect trials.

Eye movement behavior

Our pilot study aimed to verify that like increasing external cue validity (Vossel et al., 2014), increasing the predictability of an action outcome would induce a larger RS cost when this prediction was violated. This was confirmed by our pilot data. The 3 (Probability; 88/69/50%) x 2 (Validity; valid/invalid) repeated-measures ANOVA yielded a main effect of Probability ($F_{(1.7,30.8)} = 4.6$, $p = .023$, $\eta^2 = .05$, $BF_{incl} = 5.4$), reflecting overall slower responses in higher probability blocks. In terms of a main effect of Validity participants were slower for invalid compared to valid trials ($F_{(1,18)} = 31.6$, $p < .001$, $\eta^2 = .39$, $BF_{incl} > 100$), and the interaction between Probability and Validity was also significant ($F_{(1.3,24.3)} = 8.9$, $p = .003$, $\eta^2 = .048$, $BF_{incl} = 7.9$), reflecting that participants were slower in higher probability blocks in particular in invalid trials (see Figure 4). Thus, in our Pilot study, we were able to replicate the main finding by Vossel and colleagues (2014) in the context of internally generated predictions as well.

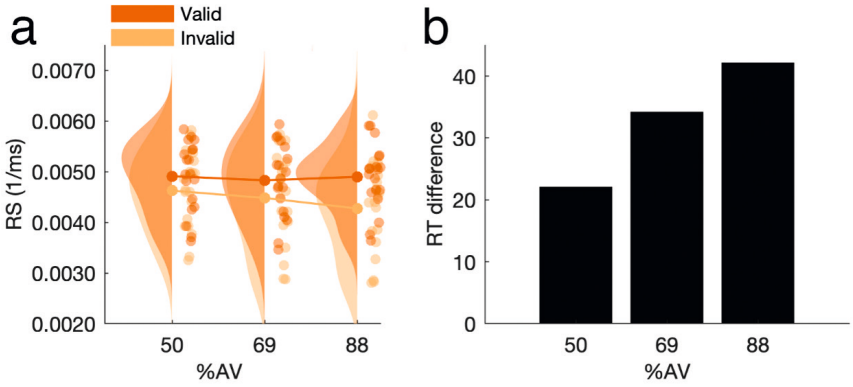


Figure 4. Eye movement results for the pilot study. Reaction speed across probability conditions split by valid and invalid trials (**a**), and the RT difference between valid and invalid trials in milliseconds (**b**). This figure shows that as the probability of a valid trial increased, the RT cost of a violated internally-generated expectation became larger.

HGF

In line with Vossel and colleagues (2014), we compared the three different response models by which the probability estimate for action validity could be transformed into RS using Bayesian model selection (Stephan et al., 2009). Rather than the precision response model, we found the belief response model to be most likely based on the protected exceedance probability (precision = .014, belief = .977, surprise = .01). The Bayes Omnibus Risk suggested all three models were not equally likely (p = .029). As can be seen in Figure 5, the trajectories for the averaged Bayesian parameters fit the presented validity evolution reasonably well. Thus, while for Vossel and colleagues (2014), the precision response model fitted best, in our case the optimal relationship between the allocation of attentional resources and predictability is a linear one. The question is whether this difference is specific to our action version of the experiment. To answer this question, we turn to our main experiment in which we directly contrast internally generated predictions through actions with externally induced predictions by cues.

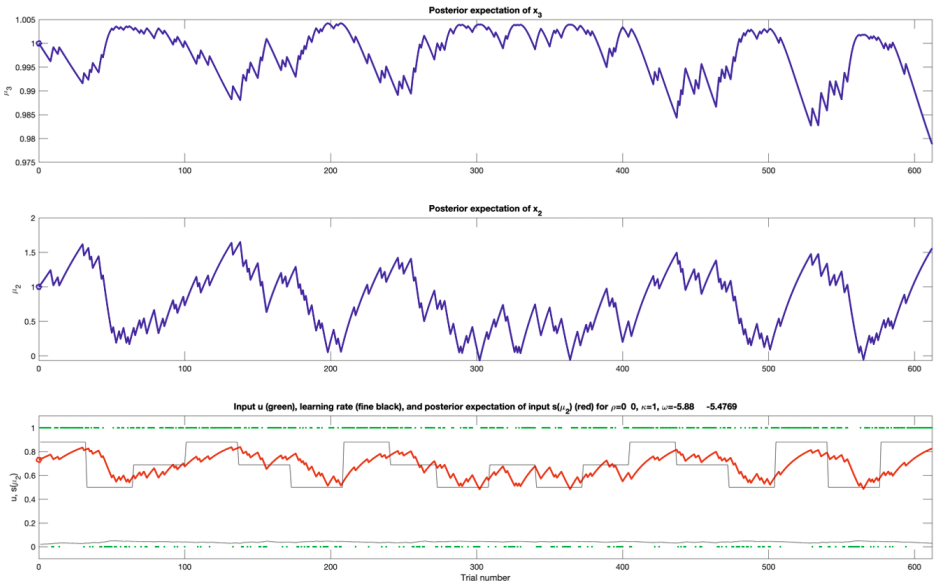


Figure 5. Trajectories for all three HGF levels derived from the averaged parameters belonging to the "belief" response model based on the Pilot data. The lowest panel shows the trial-by-trial probability estimate (red) against the actual probability blocks (black). It shows how the model successfully estimates and tracks the real probability relationship between action/cue and stimulus location.

Main experiment

Similar to Vossel et al (2014), in total 21.6% of data was excluded from further analyses. 15.8% of trials were excluded because the saccade started more than 1° from fixation, 2.6% because saccadic RT was smaller than 90 ms, 2.1% due to less than two-thirds of the distance being traversed to one of either target locations, .53% because no saccade was detected, .26% because trials consisted of more than 20% missing values, and .16% due to incorrect trials.

Eye movement behavior

We first determined whether participants were overall faster on valid trials. The 3 (Probability; 88/69/50%) x 2 (Validity; valid/invalid) x 2 (Expectation; action-/cue-induced) repeated-measures ANOVA revealed a significant main effect of Validity ($F_{(1,24)} = 59.3$, $p < .001$, $\eta^2 = .15$, $BF_{incl} > 100$), reflecting faster responses (higher RS) for valid compared to invalid trials. In addition, contrary to our preregistered prediction, participants became slower in higher probability blocks ($F_{(2,48)} = 3.2$, $p = .048$, $\eta^2 = .005$, $BF_{incl} = .08$). A third main effect of Expectation indicated that participants were overall slower in the action compared to the cue condition ($F_{(1,24)} = 6.3$, $p = .019$, $\eta^2 = .14$, $BF_{incl} > 100$).

We expected to find that expectations stemming from self-initiated actions would induce a larger reaction-time cost when violated compared to externally-cued expectations. Yet, the predicted significant three-way interaction between Probability, Validity, and Expectation was only trend level significant ($F_{(2,48)} = 2.9$, $p = .066$, $\eta^2 = .002$, $BF_{incl} = .15$) (see Figure 6). We also did not find an interaction between Validity and Expectation ($F_{(1,24)} = 2.7$, $p = .12$, $\eta^2 = .002$, $BF_{incl} = .26$). The same was true for the interaction between Probability and Validity, which was also trend-level significant ($F_{(2,48)} = 2.6$, $p = .089$, $\eta^2 = .002$, $BF_{incl} = .1$). Thus, while we found a pattern of findings in the expected direction, it was not statistically robust.

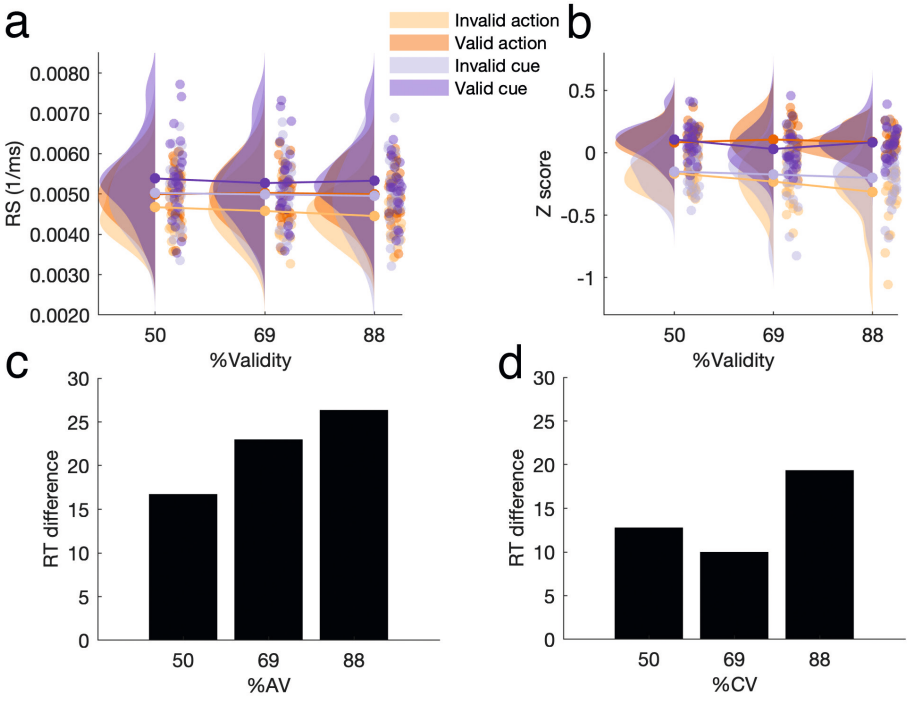


Figure 6. Eye movement results for the main experiment. Reaction speed across action/cue validity (AV/CV) conditions and split by **a**) valid and invalid trials, **b**) the same results but normalized to account for the overall difference in RS between the action and cue condition, and the reaction time difference between valid and invalid trials in milliseconds for the **c**) action and **d**) cue condition. While the RT cost for violated predictions is larger for 88% compared to 50% probability blocks in both conditions, we don't find a consistent gradual increase in the cue condition as a function of cue validity.

Control analyses

We did not expect to find an effect of session order concerning whether the action or cue condition was completed first, but this expectation did not hold up. While we did not find a between-subject effect of Order ($F_{(1.23)}$ = .34, p = .57, η^2 = .009, BF_{incl} = .56), we did find a number of interactions. The interaction between Order and Expectation indicated that participants were faster in the cue session if they completed the action session first, but if the cue condition was completed first there was no difference ($F_{(2,23)}$ = 4.7, p = .041, η^2 = .034, BF_{incl} > 100). Order interacted with Probability in that participants were slower in 88% probability blocks if they completed the action session first, while they were fastest in 50% probability blocks when the cue session was completed first. ($F_{(2,46)}$ = 6.5, p = .003, η^2 = .003, BF_{incl} = .29). We also found a three-way interaction between Probability, Expectation, and Order ($F_{(2,46)}$ = 3.7, p = .03, η^2 = .002,

BF_{incl} = .21), where participants were faster for all probabilities in the cue session only if they had already completed the action session. This difference between sessions across probabilities was absent if the cue session was completed first. Finally, we found a four-way interaction between Probability, Expectation, Validity, and Order ($F_{(2,46)}$ = 3.3, p = .045, η^2 > .001, BF_{incl} = .12), reflecting an enhanced three-way interaction between Probability, Expectation, and Validity if the action session was completed first. The task results outside Session Order reported above did not change in a meaningful way in this model. Given the relatively small sample size for this between-subject Order analysis, and the unexpected direction of some of the Order effects observed, it is unclear how to weigh these results.

In line with Vossel and colleagues (2014), but in addition to our preregistration, we conducted a control analysis for Time, in which we added a within-subject factor contrasting the first and second half of each session. This analysis yielded a main effect of Time ($F_{(1,24)}$ = 102.6, p < .001, η^2 = .25, BF_{incl} > 100), reflecting faster responses in the second half. Time interacted with Probability ($F_{(2,48)}$ = 30, p < .001, η^2 = .02, BF_{incl} = .2), to the point that participants were slower in 88% probability blocks in the first half compared to lower probability blocks, but they were faster in 88% probability blocks in the second half. Finally, we found a three-way interaction between Time, Probability, and Expectation ($F_{(2,48)}$ = 5, p = .01, η^2 = .004, BF_{incl} = .002), in that the difference in response time between the first and second half was larger for the action session, in particular for 50% and 88% probability blocks.

HGF

As to our Bayesian modelling results, we expected to replicate Vossel and colleagues (2014) in terms of the optimal response model ("precision"), in the cue as well as the action condition, although our pilot results identified the belief response model as the optimal response model for the action condition. For the action condition, Bayesian model selection indeed revealed that the precision response model was most likely (protected exceedance probability; precision = .34, belief = .32, surprise = .33), but the Bayes Omnibus Risk was sufficiently high (p = .91) to suggest that the three response models did not differ in a meaningful way. For the cue condition the belief response model came out on top (protected exceedance probability; precision = .29, belief = .48, surprise = .23), but in this case as well, the Bayes Omnibus Risk did not suggest a meaningful difference between models (p = .69).

We selected the parameters from the best model (precision for action condition, belief for cue condition) to establish whether belief updating differed

under the influence of externally generated cues versus internally generated actions. Specifically, we compared the subject-specific parameters ω and ϑ reflecting the speed of trial-wise belief updating concerning cue/action validity (ω) and the belief about the volatility of cue/action validity (ϑ). We expected participants to showcase a lower ω value in the cue condition, reflecting slower updating of predictions. Yet, contrary to our expectation a paired-samples t-test did not reveal a difference between the cue and action condition for ω_2 ($t_{(24)} = 1.1$, $p = .28$, $d = .003$, $BF_{10} = .03$). However, we did find that participants believed volatility was higher in the action condition as reflected by ω_3 ($t_{(24)} = 2.7$, $p = .01$, $d = .3$, $BF_{10} = 3.9$).

Explorative analysis

To account for the unexpected effects of (session) Order reported above, and to stick close to the original study by Vossel and colleagues (2014) in which participants only participated in one session of the cue condition, we conducted additional analyses in which we regarded the first session separately. Said differently, we regarded the first session as a between-subject study in which one group (n = 10) completed the cue session, while another group completed the action session (n = 15) to examine if we could replicate Vossel and colleagues (2014) with this more similar design.

Eye movement behavior. The 3 (Probability; 88/69/50%) x 2 (Validity; valid/invalid) x 2 (Expectation; action-/cue-induced) mixed ANOVA did reveal, as can be seen in Figure 7, the expected three-way interaction between Probability, Validity, and Expectation ($F_{(1.3,31)} = 4$, $p = .044$, $\eta^2 = .004$, $BF_{incl} = 5$), reflecting that self-generated predictions through actions incurred a larger reaction time cost in particular in the high probability condition. Moreover, a trend-level significant between-subject main effect of Expectation was observed ($F_{(1,23)} = 3.7$, $p = .068$, $\eta^2 = .11$, $BF_{incl} = 68.9$). Within-subject main effects were similar to the original analyses above: responses were slower in higher probability blocks ($F_{(1.5,35.2)} = 5$, $p = .019$, $\eta^2 = .007$, $BF_{incl} = 6.8$) and on invalid trials ($F_{(1,23)} = 46.7$, $p < .001$, $\eta^2 = .066$, $BF_{incl} > 100$). We furthermore found a significant interaction between Probability and Expectation ($F_{(1.5,35.2)} = 7.3$, $p = .004$, $\eta^2 = .001$, $BF_{incl} = 16.8$), reflecting that only in the action condition, participants became slower in higher probability blocks. The interaction between Validity and Expectation was also significant ($F_{(1,23)} = 5.4$, $p = .029$, $\eta^2 = .008$, $BF_{incl} = 47.1$), reflecting a larger reduction in response time for valid vs. invalid trials in the action condition. We did not find a significant interaction between Probability and Validity ($F_{(1.3,31)} = 1.7$, $p = .2$, $\eta^2 = .002$, $BF_{incl} = 1.4$). Thus, when we look solely at the first session, we find that violation of predictions through invalid trials

incurred a larger reaction time cost when a prediction stems from action rather than from a sensory cue, in particular under high probability conditions.

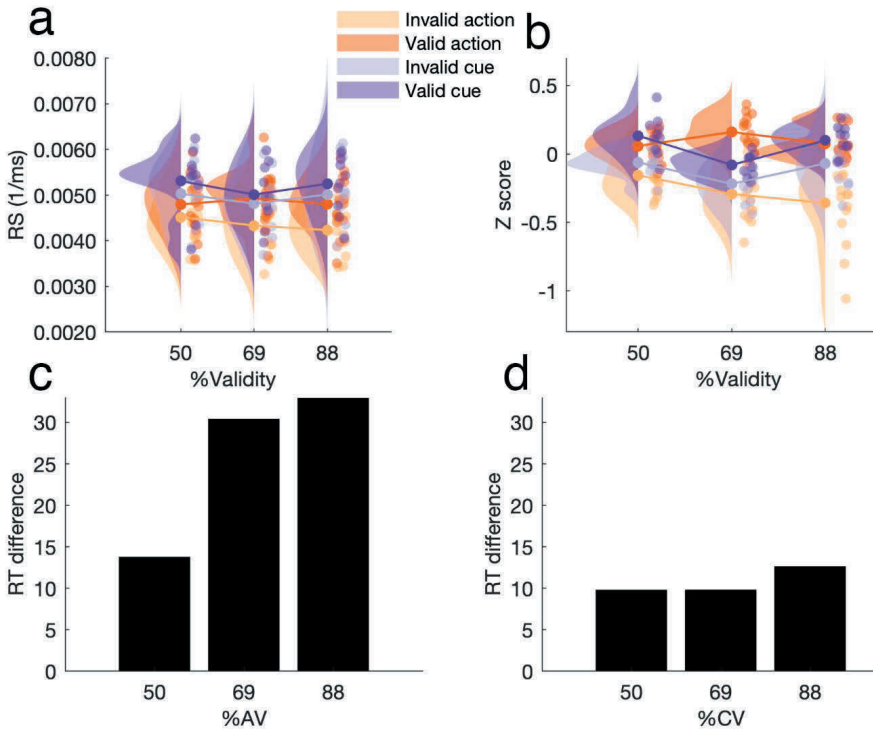


Figure 7. Eye movement results for only the first session of the main experiment. Reaction speed **a**) across probability conditions and split by valid and invalid trials, **b**) the same results but normalized, and the reaction time difference between valid and invalid trials in milliseconds for the **c**) action and **d**) cue condition. The RT cost for violated predictions is larger for 88% compared to 50% probability blocks in the action condition, but this difference is practically absent in the cue condition.

HGF. We repeated the procedure to select the best response model for the first session only. For the action condition, the surprise response model was most likely (protected exceedance probability; precision = .32, belief = .32, surprise = .35), but there was again little indication to assume the difference between these models was meaningful as reflected by the Bayes Omnibus Risk (p = .87). For the cue condition Bayesian model selection suggested the belief model was again best (protected exceedance probability; precision = .36, belief = .41, surprise = .23), but the Bayes Omnibus risk here too did not suggest a meaningful difference (p = .68)

We once again took the subject-specific parameters ω and ϑ from the winning models to compare them between the cue and action condition. An independent samples t-test did not reveal a difference for either ω_2 ($t_{(23)} = .83$, $p = .42$, $d = .34$, $BF_{10} = .48$) or ω_3 ($t_{(23)} = .25$, $p = .8$, $d = .1$, $BF_{10} = .38$). Thus, when controlling for time by reducing our analyses to only the Session 1 data, our modelling results did not change in a meaningful way.

Controlling for time. For the first session only as well, and in line with Vossel and colleagues (2014), we controlled for differences between the first and second half of the experiment by including a within-subject factor Time in our ANOVA. This resulted in a 2 (Time; first/second half) x 3 (Probability; 88/69/50%) x 2 (Validity; valid/invalid) x 2 (Expectation; action-/cue-induced) mixed ANOVA. Participants were again significantly faster in the second half of the session compared to the first half ($F_{(1,23)} = 82.4$, $p < .001$, $\eta^2 = .2$, $BF_{incl} > 100$). Time interacted with Probability ($F_{(1.8,43.1)} = 19.2$, $p < .001$, $\eta^2 = .02$, $BF_{incl} > 100$): responses became slower with higher probability blocks, but only in the first half of the session. Time did not interact with other variables (all $p > .12$). Said differently, unlike Vossel and colleagues, we still found differences between the first and second half of the experiment, but these did not differ between the action and cue groups.

5.5. Discussion

In the present study, we addressed the following question: how does the response to induced predictions and their violation differ when these predictions stem from externally induced sensory cues compared to internally generated actions? We hypothesized that self-generated predictions would incur a larger reaction time cost when violated compared to cue-induced predictions, given that self-generated predictions allow a system to engage actively in testing and revising its models instead of being at the whim of the external environment to do so. We explored this question specifically within the context of predictive processing, and under the assumption that our participants perform approximate Bayesian inference. In absolute terms we found supporting evidence for our hypothesis: participants showcased a larger reaction time difference between valid and invalid trials in the action condition, particularly for high-probability trials, but it must be noted that the main 3-way interaction we were after remained at trend level significance for our main preregistered analyses, and only became statistically significant once we exploratively isolated the first session to account for unexpectedly observed session order effects.

In terms of modelling, only in our pilot data were we able to establish a difference between response models through a significant Bayes Omnibus Risk, but even then, it was not the "precision" model that was optimal, as we expected to find based on the study by Vossel and colleagues (2014). In terms of model parameters, the difference between action and cue conditions was reflected in the volatility parameter ω_3 in our main analyses instead of the hypothesized speed of belief-updating parameter ω_2. This difference disappeared for our explorative analysis isolating only the first session, possibly due to the small groups being compared (N = 10 and N = 15). Overall, if these behavioral and modelling results are to be believed, it does seem that the larger reaction time cost induced by the violation of action-induced predictions translates to participants becoming more sensitive to the volatility of relevant probability relationships in their surroundings.

A difference between how stimuli stemming from internally generated and externally induced predictions affect model updating was expected based on the idea that self-generated predictions are generally afforded more weight compared to externally induced predictions (H. Brown et al., 2013), because they allow a system to put its own models to the test, instead of being reliant on external change in its environment. This idea aligns with the way fine-tuning of behavior is understood within the context of sensorimotor learning. As in the case of male songbirds who in a first "sensory" stage store a song-template in memory, only to put this template to the test in a second "sensorimotor" stage where they actively but gradually approach the stored template (M. S. Brainard & Doupe, 2000). In our experiment, the "template" is the intuitive relationship between a left and right action with a left and right visual effect which is put to the test repeatedly. A problem with transposing this example to the visual domain is that it is unclear whether there are anatomical connections directly connecting the motor cortex to the visual cortex, as is the case for the auditory cortex (Reznik & Mukamel, 2019). Nevertheless, it could be that such a process takes place between the motor cortex and more abstract representations such as "left" and "right" instead of purely visual representations per se. Indeed, such a proposal would fit with predictive processing accounts of sensory attenuation where neural responses to predicted stimuli are reduced for aggregated signals across many neurons, possibly reflecting a sharpening of the relevant neural representation (Bell et al., 2016; Kok et al., 2012; Press et al., 2020; Reznik & Mukamel, 2019), especially when predictions are self-generated (H. Brown et al., 2013; Korka et al., 2019, 2022). It is possible that increased sensitivity to violation for self-generated predictions comes from increased sharpening given that self-generated predictions are weighted more heavily by the system in question.

Several limitations pertain to the present study. Unlike Vossel and colleagues (2014), we found an effect of Time we did not hypothesize: participants were faster in the second half of the experiment compared to the first. One difference between our and the original study is that ours took longer to complete. Whereas Vossel and colleagues report their experiment lasting 35 minutes (p. 1442), ours took over an hour on average to complete. In the original study, the authors report allowing one and four short rest periods across their two datasets, which were increased from the first to the second to increase data quality. With this in mind, we allowed for eight short rest periods, which will have lengthened our experiment. More importantly however, participants in the study by Vossel and colleagues (2014) responded faster and made more errors (~5% vs < 1%). This difference may indicate a difference in speed-accuracy tradeoff between the studies. As such, future research should emphasize speed over accuracy in attempted replications like these. More practice trials may reduce the time effect as well by eliminating task familiarization during experimental trials.

We also found a session order effect we did not hypothesize. Participants' performance was faster in the cue session after completing the action session first, while participants were not faster in the action session after having completed the cue session first. Apparently, a task like this is learned more readily in the cue condition. This makes intuitive sense because the cue condition required less from participants. In the cue session, participants merely had to sit still and move their eyes to the target when it appeared, whereas in the action session they also had to choose an action every single trial. We chose a within-subject design to exploit the increase in statistical power and reduction in random noise these designs are privy to, but given these order effects, it may be better to opt for a between-subject design in future research. Care should be taken in recruiting enough participants given that even with our within-subject design, our effects of interest remained at trend level.

In addition to the limitations that pertain to our basic behavioral results, we were unable to replicate the modelling results by Vossel and colleagues (2014). The difference resides in the best fitting response model, which stands for the way the trial-by-trial probability of cue or action validity is transposed to response speed through the computation of the attentional factor α. In the original study this transposition was best governed by precision ($\pi^{(t)}$) in line with the proposal by Feldman and Friston (2010), but in our case only for our pilot data did we find a clear winning model ("belief"). While the behavioral limitations discussed above may have contributed to these ambiguous modelling results, there are additional angles to consider. Whereas extensive

model comparison was performed in the original study (Vossel, Mathys, et al., 2014), in subsequent studies employing the same task an optimal "precision" response model is assumed without repeated model comparisons for these new datasets (Vossel, Bauer, et al., 2014; Vossel et al., 2015), while in other work with different attention tasks, it was not the "precision" but the "belief" model that fitted best (Dombert et al., 2016; Kuhns et al., 2017). It's possible that the tasks used were sufficiently different to warrant a different optimal response model, but it is also possible that the model fitting procedure is not as robust as we may desire. The Vossel et al. (2014) study had an even smaller sample size than the current study, which may have led to less robust results. Future research would do well to elucidate under what conditions different response models fit best with larger samples sizes. Perhaps a better modelling approach altogether to test the current research question would be one in which the difference is considered between perceptual and active inference, as they've been theorized to optimize two separate free energy functionals (Parr & Friston, 2019).

While the present study leaves room for improvement in experimental design, the results we report nevertheless uphold the importance of considering the difference between prediction of self-generated and externally imposed stimuli, especially within the context of predictive processing. While both perception and action serve the same goal: the minimization of prediction error or free energy (Clark, 2013; Friston, 2010; Hohwy, 2013), it seems likely that self-generated predictions are afforded more weight by the generating system. It may be worth considering letting go of some experimental control for the sake of an inside-out perspective.

Acknowledgements
We would like to thank Simone Vossel for helpful modelling guidance.

Chapter 6

Summary and general discussion

The present dissertation took the free-energy principle (FEP) as its starting point, from which we tried to draw both philosophical and empirical consequences. The attempt to depart from a new starting point comes with a dilemma. The disadvantage of doing something entirely new is that it may be hard to do justice to and link up with previous research. When we do decide to link up with what's done before, we run the risk of inadvertently remaining stuck in the old point of departure. For our empirical work presented in **chapter 2 and 3** we nevertheless opted for the latter, presenting stimuli to our participants to which they had to respond. This allowed us to determine the role of the basal ganglia and its irrigation by dopamine in the process of perceptual inference.

Both **chapter 2 and 3** departed from the idea that conscious perception depends on global amplification of sensory input, and that the basal ganglia (BG) and its irrigation by dopamine play a crucial role in gating information, conscious access, and the selection of a relevant internal model given available sensory data. The BG are thought to play this role due to their modulatory influence on thalamocortical connectivity. Because much of the evidence implicating the BG in these processes in humans is correlational, we explored two ways of manipulating BG activity experimentally.

Chapter 2 describes a double-blind crossover pharmacological study in which we administered cabergoline – a dopamine D2 agonist – and placebo to 30 healthy participants. Dopamine D2 receptors are abundant in the basal ganglia, specifically in the striatum. We reasoned that striatal dopamine may influence the contents of consciousness by controlling which 'internal model' or one interpretation of the current sensory state dominates perceptual experience. Under both the drug and placebo condition, we subjected participants to several well-established experimental conscious-perception paradigms, such as backward masking and the attentional blink task. We found no evidence in support of an effect of cabergoline on conscious perception: key behavioral and event-related potential (ERP) findings associated with each of these tasks were unaffected by cabergoline. Our results cast doubt on a causal role for dopamine in visual perception. It remains an open possibility that dopamine has causal effects in other tasks, perhaps where perceptual uncertainty is more prominent, or where participants are actively involved in generating their percepts.

Chapter 3 reports on a preliminary investigation in four patients to explore whether deep brain stimulation (DBS) in the BG might improve conscious perception. In our study, treatment-resistant obsessive-compulsive disorder (OCD) patients with a striatal DBS implant completed two canonical conscious perception tasks: emotion-induced blindness and backward masking. We

found preliminary evidence in support of a role played by the BG in conscious perception at the behavioral level: patients performed better when stimulation was active, but we could not establish neural effects corresponding to these behavioral findings, possibly due to our small sample size.

Chapter 4 investigates the philosophical heritage implicitly touched on by the FEP, which provides an alternative philosophical and historical background for present-day research in cognitive neuroscience. Friston's FEP has been received with great enthusiasm. With good reason: it not only makes the bold claim to a unifying theory of the brain, but it is presented as an a priori principle applicable to living systems in general. In this paper, we set out to show how the breadth of scope of Friston's framework converges with the dialectics of Georg Hegel. Through an appeal to the work of Catherine Malabou, we aimed to demonstrate how Friston not only reinvigorates Hegelian dialectics from the perspective of neuroscience, but that the implicit alignment with Hegel necessitates a reading of the FEP from the perspective of Hegel's speculative philosophy. It is this reading that moves beyond the discussion between cognitivism and enactivism surrounding Friston's framework; beyond the question whether the organism is a secluded entity separated from its surroundings, or whether it is a dynamical system characterized by perpetual openness and mutual exchange. From a Hegelian perspective, it is the tension between both positions itself that is operative at the level of the organism; as a contradiction the organism sustains over the course of its life. Not only does the organism's secluded existence depend on a perpetual relation with its surroundings, but the condition for there to be such a relation is the existence of a secluded entity. We intended to show how this contradiction – tension internalized – is at the center of Friston's anticipatory organism; how it is this contradiction that grounds the perpetual process of free energy minimization.

Chapter 5 is the report of a study attempting to contrast the FEP's perspective with that of traditional cognitive neuroscience. While the FEP casts the brain as an organism's predictive model of how its world works and will continue to work in the future in which action is afforded a central place, research on the brain's predictive capacities remains beholden to traditional research practices in which participants are passively shown stimuli without their active involvement (as we also did in Chapters 2 and 3). The current study is an investigation into ways in which self-generated predictions may differ from externally induced predictions. Participants completed a volatile spatial attention task under both conditions (externally/cue-induced, internally/action-induced) on different days. We used the Hierarchical Gaussian Filter,

an approximate Bayesian inference model, to determine subject-specific parameters of belief-updating and inferred volatility. We found preliminary evidence in support of self-generated predictions incurring a larger reaction time cost when violated compared to predictions induced by sensory cue, which translated to participants' increased sensitivity to changes in environmental volatility. Our results suggest that internally generated predictions may be afforded more weight, but these results are complicated by session order and duration effects, as well as a lack of statistical power.

The empirical studies reported in this dissertation can be divided into two themes: the role of the BG in conscious perception (**chapter 2 and 3**), and the role of action in perception and belief updating (**chapter 5**). What these three chapters share is a lack of clear evidence in support of the hypotheses under investigation. The (practical) reasons for these inconclusive results are multiple and can be found in each respective chapter. These results could be taken as lack of support for the FEP, and by extension as a reason to abandon the FEP altogether. In my view that would be a mistake. For one, chapters 2 and 3 still took a traditional approach to the study of conscious perception: participants were presented with stimuli on a computer screen to which they had to respond. This left little room for effects of participant's active involvement in the processes under study. Moreover, all chapters focused on "cold" cognitive measures, which may not account for the affectively charged information organisms use to orient themselves in the world. In the remainder of this section, I would like to consider briefly the possibility that up until now, we have not thought through the theoretical implications of the FEP far enough, and that by extension we do not yet have an adequate theoretical orientation to base our empirical studies on. Instead of reverting away from the FEP back to the traditional perspective, perhaps we need to take another step further away from it.

What makes the FEP an appealing framework for cognitive neuroscience is that it offers an incredibly simple conception of biological systems capturing little more than what a system must do to keep existing; namely, constrain the actual states it occupies out of all possible states, or what comes down to the same imperative: minimize long-term average surprise (or prediction error). Crucial is that this idea is grounded in modern physics, instead of the 17th century empiricist philosophy of the traditional view.

Broadly speaking, the traditional empiricist view – and by extension a large portion of cognitive neuroscience research – is focused on (visual) perception directed at the world outside the perceiving brain in question. Up until now, the

FEP has been employed mainly as a justification to shift away from an emphasis on perception to action (cf. Buzsáki, 2019). The argument to do so is that from the perspective espoused by the FEP, perception merely serves to inform what an organism must do to survive. As justifiable as this shift is, what it lacks is the information needed by the system to inform its actions to stay alive. Such information is not a matter of perception, but of affect. It is through affect that a system registers it is departing or has departed from its preferred states in proximity of homeostasis. Emphasis on affect would allow us to start out from a system's internal state in terms of what it lacks, such as food, rest, information, or whatever else as the main driver of action, which perception facilitates. In addition to a shift from perception to action, perhaps what we need to fully draw out the FEP's implications is a shift from action to affect.

In his 2021 book *The Hidden Spring*, the neuroscientist and psychoanalyst Mark Solms employs the FEP in search of ways out of the empiricist view in relation to how the brain instantiates consciousness:

> "The everyday observation that our consciousness consists mainly of perceptual images of events going on around us suggests that consciousness flows in through the senses. This common-sense view has no doubt been with us since we first began to think about such matters. In the seventeenth and eighteenth centuries, it gave rise to the 'empiricist' philosophies of John Locke and David Hume. They theorised that the mind – which begins as a blank slate – acquires all its specific characteristics from impressions left by sensory vibrations. The impressions were supposed to become associated with each other through regular conjunctions of various kinds to produce our memory images of objects, which in turn became the basic building blocks of more abstract ideas." (p. 60-61)

Solms' alternative to the empiricist view not only includes an inside-out perspective (cf. Buzsáki, 2019), but he leverages affect to depart from the dominant view on how to study consciousness. For Solms, an organism is conscious when it registers its drives, where drives are understood as demands of the body on the mind for work. Examples include the simple drive to resolve hunger, uncertainty, correcting to a preferred state of thermoregulation, avoiding tissue damage (pain), and fatigue, with the possibility of more complex emotional drives in more complex organisms like humans. Solms concludes: "Consciousness is endogenously generated; all of it. Consciousness

at its source is affect. Then it is extended outwards onto perception, to evaluate perceptual inferences [...].'' (p. 189). Casting consciousness as affective stands in stark contrast to the perceptual and cognitive theories that predominate consciousness research (Hohwy & Seth, 2020; Seth & Bayne, 2022).

Perhaps the main limitation of each of our empirical studies – resulting in our lack of results – is not that we tried to start out from the FEP, but that we still adhered all too closely to the traditional empiricist view, by resorting to passive visual perception of arbitrary stimuli as our main process under investigation. Alternative approaches abound from the perspective of the FEP, for it raises action, proprioceptive, and interoceptive (affective) inference to the forefront of biological functioning (Allen et al., 2022; Azzalini et al., 2019; Pezzulo et al., 2015; Seth & Friston, 2016), alongside passive exteroceptive inference (perception).

A promising avenue for future research in this regard comes from active inference, a mathematical derivative of the FEP (Parr et al., 2022; Sajid et al., 2021). Active inference is a computational framework not only allowing for modelling of behavioral data, but also for the construction of virtual agents leveraging affect to motivate the satisfaction of needs and thereby to maintain approximate homeostasis. It thereby enters into direct competition with reinforcement learning (RL) schemes that predominate the field of artificial intelligence (Barto et al., 2013; Sutton & Barto, 1998). In contrast to RL, active inference provides a Bayesian mechanics from which arguably all normative optimization schemes can be derived as special cases. For example, active inference reduces to RL if we ignore uncertainty about hidden states (Friston et al., 2021). This combination of capturing behavioral modelling and the construction of virtual agents under the same first-principled umbrella opens up the possibility of parallelizing the study of active inference in virtual and actual biological systems. For example by constructing an environment for virtual agents which can be transposed to a virtual reality setting for actual biological systems (human or animal)(Anggraini et al., 2018). The combination of computational rigor together with the possibilities of virtual world-building may be what the field of cognitive neuroscience needs not only to shed its empiricist baggage, but to turn it into a field capable of doing justice to what human brains and bodies are capable of.

Chapter 7

Nederlandse samenvatting

Hoe verwerkt het brein informatie? Wat doet het brein? Terwijl beide vragen in dezelfde richting wijzen, stelt het verschil tussen beiden het thema scherp waar het in deze dissertatie mede om gaat. Het beeld van het brein als passieve verwerker van informatie staat onder druk. De hedendaagse neurowetenschap biedt een nieuw beeld. Met het vrije-energie principe (VEP) [*free-energy principle*] van de Britse psychiater en natuurkundige Karl Friston verschijnt het brein als een habitueel verwachtingsmodel dat actief deelneemt in haar eigen vormgeving: dat doet. In plaats van een orgaan geketend aan invloeden van buiten is het brein een structuur die uitstaat naar buiten en naar de toekomst, op basis van zowel een evolutionair, cultureel als individueel verleden. Het belang van dit nieuwe breinbeeld is tweeledig.

Minimale vrije-energie minimalisatie

Ten eerste biedt het een raamwerk van waaruit we een begin kunnen maken biologische systemen in hun geheel te begrijpen en wetenschappelijk te bestuderen. Als we het vrije-energie principe terugbrengen tot de essentie, dan staan drie hoofdpunten centraal. Het VEP vertrekt vanuit de aanname dat een biologisch systeem een begrenzing in stand houdt tussen zichzelf en de buitenwereld. Doet het dat niet, dan zou een dergelijk systeem spoedig opgaan in zijn omgeving en daarmee ophouden te bestaan. Anders gezegd: geen minimale begrenzing, geen biologisch systeem.

Ten tweede houdt een biologisch systeem een dergelijke begrenzing in stand door ervoor te zorgen dat het een beperkt aantal toestanden inneemt uit alle mogelijke toestanden die het mogelijkerwijs in kan nemen. Ons lichaam doet niet anders gedurende de dag door zorg te dragen dat zaken zoals lichaamstemperatuur en bloedsuikerspiegel niet buiten bepaalde grenzen raken. Als ons biologisch systeem hier niet in zou slagen, dan zouden we spoedig sterven. Ditzelfde proces kunnen we generaliseren naar processen die minder noodzakelijk zijn voor ons directe voortbestaan, zoals processen gebaseerd op voorkeuren wat betreft informatievoorziening en emotionele behoeften. Zo zijn sommige mensen bijvoorbeeld toleranter wat betreft onzekerheid dan anderen.

Het derde hoofdpunt betreft Bayesiaanse statistiek en in het bijzonder de manier waarop de beperking van in te nemen mogelijke toestanden onder bepaalde voorwaarden wiskundig gezien equivalent is aan het maximaliseren van bewijs voor een Bayesiaans model. Het idee is dat biologische systemen al dan niet impliciet een vorm van Bayesiaanse gevolgtrekking [*inference*] voltrekken. Deze vorm van statistiek combineert de a priori waarschijnlijkheid van nieuwe data en de aannemelijkheid [*likelihood*] van deze data *gegeven*

een model, om te komen tot de a posteriori waarschijnlijkheid van de data in kwestie. Of het brein Bayesiaanse gevolgtrekking belichaamt is vooralsnog een open vraag die meer onderzoek vergt, maar een mogelijke werking is dat het receptief veld van neuronen – de data waar neuronen op vuren – neerkomt op de waarschijnlijkheid van de data waar ze aan blootgesteld worden. Het VEP stelt dat het brein niet een model van haar omgeving heeft, maar een model van haar wereld *is*. Dit betekent dat eerdere ervaringen en het verleden inherent verbonden is met de manier waarop het brein waarneemt en handelt.

Vrije energie is de kwantiteit die het biologisch systeem in haar geheel heeft te minimaliseren om te kunnen voorspellen hoe haar omgeving werkt en verandert, en uiteindelijk om te blijven bestaan. Onder bepaalde aannames is de minimalisatie van vrije energie gelijk aan de minimalisatie van voorspellingsfout [*prediction error*]: het verschil tussen de voorspellingen die voortkomen uit het neurale Bayesiaanse model en de data waar het model zich geconfronteerd mee ziet. Er zijn meerdere manieren waarop het systeem poogt dit verschil te minimaliseren. De eerste manier is via waarneming, door deze aan te passen aan de nieuwe data. De tweede manier is via actie, door te handelen en de data in kwestie te veranderen. Stel je wandelt buiten op een mistige ochtend en je ziet iets in de verte zonder precies te kunnen uitmaken wat je ziet. Een mogelijkheid is te concluderen dat het wel een verkeersbord moet zijn (waarneming). Een andere mogelijkheid is om te handelen door je dichterbij datgene wat je ziet te begeven en zo je onzekerheid te verminderen. Waarneming en actie zijn twee manieren om vrije energie te minimaliseren aan de zintuigelijke randen van het systeem. Wat het systeem ook doet is het model offline optimaliseren door te dromen en daarmee mogelijke scenario's te simuleren en uit te werken. Online is er ook de mogelijkheid van de verbeelding, door in wakkere toestand mogelijke scenario's te doordenken. De mate waarin een biologisch systeem in staat is tot verbeelding hangt af van hoe complex het biologische systeem in kwestie is. Mensen zijn wat dat betreft tot meer in staat dan eencelligen.

Deze drie hoofdpunten tezamen maken van het VEP een op de moderne natuurkunde geïnformeerd startpunt om na te denken over biologische systemen in het algemeen, en de hersenen in het bijzonder. Bezien vanuit het VEP is het een brein een voorspellend model dat ons in staat stelt onszelf te begeven in de wereld om ons heen op basis van ons individuele en gemeenschappelijke verleden. Dergelijke systemen onderhouden een eigen interne dynamiek op basis waarvan ze opereren, in de voortdurende poging het model dat ze zijn af te stemmen op een omgeving die al dan niet verandert.

Empirische explorations

Een radicaal nieuw vertrekpunt zoals de VEP gaat gepaard met een dilemma wat betreft nieuw empirisch onderzoek dat dit vertrekpunt tot uitgangspunt wil nemen. Doe je iets volstrekt nieuws of probeer je aansluiting te vinden bij eerder onderzoek, ondanks dat dit eerdere onderzoek een ander vertrekpunt had? Het nadeel van iets volstrekt nieuws is dat het moeilijk is om in gesprek te gaan met en recht te doen aan eerder onderzoek. Aansluiten bij eerder onderzoek brengt het risico met zich mee dat je onbedoeld blijft hangen in het oude vertrekpunt. Desondanks hebben we in de empirische **hoofdstukken 2, 3 en 5** gekozen voor het laatste.

Bewuste waarneming

Zowel **hoofdstuk 2 als 3** gingen uit van het idee dat bewuste waarneming afhangt van globale versterking van zintuiglijke input, en dat de basale ganglia (BG) een cruciale rol spelen in het verheffen van deze informatie tot bewustzijn, oftewel in de selectie van een relevant intern model gegeven de beschikbare zintuiglijke data. Men denkt dat de BG deze rol vervullen door hun modulerende invloed op de thalamocorticale connectiviteit. Omdat veel van het bewijsmateriaal dat de BG bij deze processen betrekt correlationeel van aard is, met name bij de bestudering van mensen, hebben we twee manieren onderzocht om BG-activiteit experimenteel te manipuleren.

Hoofdstuk 2 beschrijft een dubbelblinde cross-over farmacologische studie waarin we cabergoline - een dopamine D2 agonist - en placebo toedienden aan 30 gezonde jongvolwassenen. Onder beide condities onderwierpen we deze deelnemers aan verschillende welbekende experimentele taken gericht op de bestudering van bewuste perceptie, zoals achterwaartse maskering [*backward masking*] en de aandachts-knipper [*attentional blink*] taak. Wij vonden geen bewijs voor een effect van cabergoline op bewuste waarneming: de belangrijkste gedragsbevindingen en elektro-encefalografie bevindingen voor elk van deze taken werden niet beïnvloed door cabergoline. Onze resultaten doen twijfel rijzen over een oorzakelijke rol van dopamine in visuele waarneming. Het zou kunnen dat cabergoline niet het juiste middel is om een dergelijke rol aan te tonen. Ook is het een mogelijkheid dat het causale effect van dopamine in de waarneming is weggelegd voor taken waarin perceptuele onzekerheid prominenter aanwezig is of het brein zelf zijn input genereert door middel van actie.

Hoofdstuk 3 doet verslag van een preliminair onderzoek met een kleine groep van vier deelnemers om te onderzoeken of diepe hersenstimulatie [*deep brain stimulation*] (DBS) in de BG bewuste waarneming verbetert. In

deze studie voerden behandelingsresistente obsessieve-compulsieve stoornis (OCD) patiënten met een striatale DBS-implantaat twee canonieke taken uit op het gebied van bewuste waarneming: emotie-geïnduceerde blindheid en achterwaartse maskering. We vonden preliminair bewijs voor een rol van de basale ganglia in bewuste waarneming op gedragsniveau: patiënten presteerden beter wanneer de stimulatie aanstond, maar we konden geen neurale effecten vaststellen die overeenkwamen met deze gedragsbevindingen, mogelijk door onze kleine steekproefgrootte. Aan het eind van het hoofdstuk bespreken we mogelijke implicaties en beperkingen van onze studie en schetsen we wegen voor toekomstig onderzoek.

Waarnemen naar handelen

Hoofdstuk 5 is het verslag van een onderzoek waarin we hebben getracht het perspectief van het FEP af te zetten tegen dat van de traditionele cognitieve neurowetenschappen. Terwijl de FEP de hersenen beschouwt als het voorspellende model van een organisme met betrekking tot hoe zijn wereld werkt en zal blijven werken in de toekomst waarbinnen actie een centrale plaats krijgt toebedeeld, blijft onderzoek naar de voorspellende capaciteiten van de hersenen gebonden aan traditionele onderzoekspraktijken waarin deelnemers passief stimuli te zien krijgen zonder hun actieve betrokkenheid. Deze studie is een onderzoek naar manieren waarop zelfgegenereerde voorspellingen verschillen van extern geïnduceerde voorspellingen. Deelnemers voltooiden een volatiele ruimtelijke aandachtstaak onder beide condities op verschillende dagen. We gebruikten een Bayesiaans inferentiemodel om deelnemer-specifieke parameters van belief-updating en ingeschatte volatiliteit te bepalen. We vonden enig bewijs dat de schending van zelfgegenereerde voorspellingen meer reactietijd kost in vergelijking met de schending van voorspellingen die door een sensorische cue worden opgewekt. Dit verschil vertaalde zich naar een verhoogde gevoeligheid van de deelnemers voor veranderingen in omgevingsvolatiliteit. Onze resultaten suggereren dat intern gegenereerde voorspellingen meer gewicht krijgen, maar deze resultaten worden gecompliceerd door sessievolgorde- en duureffecten, alsmede een gebrek aan statistische power. In Hoofdstuk 5 bespreken we de beperkingen van onze studie die ons verhinderden eerder onderzoek te repliceren, en manieren om deze tekortkomingen te verhelpen in toekomstige studies.

Filosofische verkenning

Naast de nieuwe empirische invalshoeken schuilt het belang van het VEP in de mogelijkheid om de dynamiek van het brein te doordenken in lijn met de

omwenteling in de moderne filosofie sinds Kant. Zoals met Kant zintuiglijkheid onderhevig is aan de a-prioristische vormen van het verstand, zo vertrekt het nieuwe brein vanuit a-prioristische verwachtingsmodellen. Dit is substantiële vooruitgang ten opzichte van de empiristische inslag (het brein als passieve informatieverwerker) die het onderzoeksveld van de cognitieve neurowetenschap tot de op dag van vandaag kenmerkt. De vraag is echter of Kantiaanse coördinaten ver genoeg gaan om de omwenteling van het VEP recht te doen. **Hoofdstuk 4** is een poging te laten zien hoe het VEP aansluit bij Hegels natuurfilosofie.

De aanzet van Hegels filosofie is om waarheid te vatten: niet alleen als substantie, maar ook als subject. Een manier waarop we deze beroemde frase uit de voorrede van de *Fenomenologie van de geest* kunnen interpreteren is dat we datgene wat bestaat niet alleen in haar objectiviteit dienen te begrijpen, maar dat we tevens recht te doen hebben aan subjectiviteit *als onderdeel van de objectieve werkelijkheid*. Het gaat om de manier waarop subjectiviteit zelf een objectief fenomeen is, zonder het subjectieve te reduceren tot louter objectiviteit. Anders gezegd, het gaat om de *objectiviteit van subjectiviteit*; In zekere zin kunnen we zeggen dat Hegels gehele filosofie erop is toegespitst om het idee van het objectief bestaande en denkende subject op de spits te drijven.

Hiervoor dienen we af te dalen in de abstracte diepten van ons voorspellend brein om niet alleen klem te geraken in abstracte bepalingen die consistente theorie bemoeilijken, maar om uitgerekend dit klemzitten te verheffen tot theorie, in zoverre dat sommige theoretische inconsistenties en tegenstrijdigheden niet slechts een tekortkoming van het denken zijn, maar iets zeggen over datgene wat we pogen te doordenken. Het frappante is dat uitgerekend Fristons VEP een beroep op de objectiviteit van subjectieve inconsistenties niet uitsluit, maar impliciet aanmoedigt. Tenslotte, in zoverre dat het voorspellend brein een afspiegeling is van de wereld waarin wij leven, zou het dan niet zo kunnen zijn dat de inconsistenties van dit voorspellend model niet alleen tekortkomingen zijn van het model – van het individuele brein in kwestie, maar in sommige gevallen daadwerkelijk iets zeggen over de wereld waarin wij leven; over de wereld die we pogen te bestuderen?

Nu kan het natuurlijk niet zo zijn dat elke tegenstrijdigheid waar we in ons denken tegenaan lopen blijk geeft van waarheid met betrekking tot datgene wat we overdenken. Soms zitten we er gewoon gruwelijk naast en schiet ons denkvermogen tekort. Het punt is dat wanneer we grondig zijn in ons denken, we zo nu en dan stuiten op onoplosbare tegenstrijdigheden, en dat die wellicht informatief zijn voor de werking van het object dat we proberen te begrijpen. De inzet van **Hoofdstuk 4** is om gebruik te maken van het grondige denken van

anderen rond zowel de open- en geslotenheid als de (on)afhankelijkheid van biologische systemen onderhevig aan het VEP, in een poging de voortslepende discussie rond deze tegenstrijdigheden op te laten gaan in een tegenstelling waar het systeem zelf onderhevig aan is. Deze tegenstelling is de dialectiek van vrije-energie minimalisatie, waar het hoofdstuk zijn titel aan ontleent.

Wetenschap voor Hegel omvat naast het uiterst serieus nemen van empirische wetenschap tevens het toestaan en de mobilisatie van dergelijke denkbewegingen in onze zoektocht naar waarheid. In aanvulling op de empirische hoofdstukken en de oplosbare inconsistenties die deze hoofdstukken kenmerken, is de inclusie van een hoofdstuk toegewijd aan de onoplosbare inconsistenties van ons biologisch systeem wat deze dissertatie maakt tot wetenschap in de meest omvattende zin.

Authorship specifications

Chapter 2: Boonstra, E. A., van Schouwenburg, M. R., Seth, A. K., Bauer, M., Zantvoord, J. B., Kemper, E. M., Lansink, C. S., & Slagter, H. A. (2020). Conscious perception and the modulatory role of dopamine: No effect of the dopamine D2 agonist cabergoline on visual masking, the attentional blink, and probabilistic discrimination. *Psychopharmacology*, *237*(9), 2855–2872. https://doi.org/10.1007/s00213-020-05579-9

E.A.B.: Methodology, Formal Analysis, Writing—original draft, Writing—review & editing, Visualization. M.R.S.: Conceptualization, Methodology, Writing—review & editing, A.K.S.: Conceptualization, Methodology, Writing—review & editing, Funding acquisition. M.B: Methodology, Writing—review & editing. J.B.Z.: Methodology, Psychiatric consultation, Writing—review & editing, E.M.K.: Methodology, Pharmacy services, Writing—review & editing, C.S.L.: Conceptualization, Methodology, Writing—review & editing, Funding acquisition. H.A.S.: Conceptualization, Methodology, Writing—review & editing, Supervision, Funding acquisition.

Chapter 3: Boonstra E. A., Bais, M. N., van Schouwenburg M. R., van den Munckhof, P., Smit, D. J. A., Denys, D., & Slagter, H.A. (in prep.). Conscious perception and the role of the basal ganglia: preliminary findings from a deep brain stimulation study.

E.A.B.: Conceptualization, Methodology, Formal Analysis, Writing—original draft, Writing—review & editing, Visualization. M.N.B.: Conceptualization, Methodology, Writing—review & editing. M.R.S.: Conceptualization, Methodology, Writing—review & editing. P.M: Surgery, Methodology, Writing—review & editing. D.J.A.S.: Methodology, Writing—review & editing. D.D: Conceptualization, Methodology, Writing—review & editing, Supervision, Funding acquisition. H.A.S.: Conceptualization, Methodology, Writing—review & editing, Supervision, Funding acquisition.

Chapter 4: Boonstra, E. A., & Slagter, H. A. (2019). The dialectics of free energy minimization. *Frontiers in Systems Neuroscience*, *13*, 42. https://doi.org/10.3389/fnsys.2019.00042

E.A.B.: Conceptualization, Writing—original draft. Writing—review & editing, Visualization. H.A.S.: Conceptualization, Writing—review & editing, Supervision, Funding acquisition.

Chapter 5: Boonstra, E. A. & Slagter, H. A. (in prep.). Learning to predict based on self- versus externally induced prediction violations: a direct comparison using a Bayesian inference modelling approach.

E.A.B.: Conceptualization, Methodology, Formal Analysis, Writing—original draft. Writing—review & editing, Visualization. H.A.S.: Conceptualization, Methodology, Writing—review & editing, Supervision, Funding acquisition.

Dankwoord / Acknowledgements

Mijn dank gaat allereerst uit naar mijn promotor en begeleider. Heleen, ik kan er nog altijd niet helemaal bij hoe jij als onderzoeker ongebreidelde kunde en grondigheid weet te combineren met niet alleen een ontstellende reikwijdte aan onderzoeksrichtingen waarin je die kwaliteiten tentoonspreidt, maar ook met onverminderde empathie voor zaken die niet direct betrekking hebben op het academische bedrijf. Helemaal in een tijd waarin zowel in mijn directe omgeving als in het nieuws constant sprake lijkt te zijn van machtsmisbruik door hoogleraren die hun werknemers erbarmelijk behandelen, geeft het zoveel hoop dat de universiteit blijkbaar nog altijd in staat is om ook mensen zoals jij te koesteren. Ik ben zo dankbaar dat ik mijn dissertatie onder jouw begeleiding heb mogen afronden. Mijn toekomst binnen de academische wereld is vooralsnog onzeker, maar mocht die toekomst er zijn, dan weet ik zeker dat de enige manier waarop ik een academisch onderzoeker wil zijn in lijn is met het goede voorbeeld dat jij me hebt gegeven de afgelopen jaren.

Mijn dank gaat ook uit naar mijn copromotor Chris O. en tweede begeleider Martine. Martine, bedankt voor onze intensieve samenwerking in het eerste anderhalf jaar van mijn PhD. Jij hebt het schrik aanjagende begin van een PhD zoveel makkelijker gemaakt met zowel je kundigheid als steun. Met enige schaamte moet ik bekennen dat ik je aanmoediging bij je afscheid om Amsterdam te verkennen nog altijd niet in de praktijk heb gebracht, maar ik waardeer de aansporing. Chris O., we mogen dan niet bijzonder intensief samengewerkt hebben, ik waardeer het enorm dat je bereid was en bent mijn copromotor te zijn.

I am grateful to the colleagues I've had the pleasure of sharing an office with at the VU (Elle, Inês, Caterina, Chris J., Eduard, and Katja). Thank you for the warm and wholesome space you created, making me feel so very much at home. More broadly, I am grateful to my colleagues outside of our office in the department of Experimental and Applied Psychology, as well as in the Brain & Cognition department at the UvA. I will miss the pantry lunches and department activities. With special thanks to Suzan. You not only gave me such a warm welcome and thoughtful tour of the VU when I arrived, but I appreciate you preventing me from getting lost in university bureaucracy. Finally, I want to extend my special thanks to Elle for being such a stable factor in our office, in general but especially during the covid lockdowns, and to Changrun and Elle for the daily lunches during the worst of the pandemic: you made lockdowns so much better.

I want to thank my fellow lab members in Heleen's Plasticity & Cognition lab (Ruben, Caterina, Chris J., Martine, Fleur, Lynn, Leon, Dirk, Josipa, Valentina, and Jim). Thank you for the rounds of updates, discussions, and feedback during our lab meetings, as well as for the sense of community you provided.

I would like to express my deepest appreciation to the members of my defense committee: prof.dr. Jeroen Smeets, prof.dr. Birte Forstmann, prof. dr. Floris de Lange, prof.dr. Clare Press, dr. Julian Kiverstein, and dr. Filip van Opstal. Thank you for spending your time on critically evaluating this dissertation, both in the lead-up to and during my defense.

I am deeply indebted to my partner Mieke. I am so glad we reconnected when we did. You're my favorite agent of chaos, expanding what I considered possible in virtually all areas of my life. Thank you for sticking with me through what was undoubtedly the most difficult year of my PhD. Whether we worked from an Amstelveen hotbox, The Scholar Guide headquarters, or a Constantia mansion, it was good because it was with you.

Ik ben dankbaar voor mijn zus Hanneloes. Loes, ik moet er niet aan denken hoe de afgelopen jaren waren geweest zonder jou. Zelfs de meest moeilijke zaken voelden dragelijk omdat we ze samendeden.

Ik betwijfel of deze dissertatie de eindstreep had gehaald zonder de niet aflatende steun van vrienden en de mogelijkheid om tijd met jullie door te brengen: Alex, Juliette, Wendy, Tycho, Kitty, Ronald, Chris H., Stefan, Masja, Peter, Washant, Sharon en Sander.

Laila, our relationship may not have lasted the full duration of our PhDs, but I feel incredibly grateful for all the walks, burgers, and wraps we shared both during the weekends in Salzburg and Amsterdam, as well as during the first covid lockdown.

Als laatste wil ik mijn ouders noemen. Frank, bedankt voor de manier waarop je het belang van zowel wetenschappelijk onderzoek als kritiek op de denkbeelden die we met ons meedragen weet te benadrukken. Annemieke, bedankt voor het ogenschijnlijk eindeloze aantal zondagavonden waarop we een maaltijd, een kop thee en al dan niet een potje rummikub hebben gedeeld.

31 oktober 2022
Evert Boonstra

References

Acevedo, N., Bosanac, P., Pikoos, T., Rossell, S., & Castle, D. (2021). Therapeutic Neurostimulation in Obsessive-Compulsive and Related Disorders: A Systematic Review. *Brain Sciences*, *11*(7), 948. https://doi.org/10.3390/brainsci11070948

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, *218*(3), 611–643. https://doi.org/10.1007/s00429-012-0475-5

Afrasiabi, M., Redinbaugh, M. J., Phillips, J. M., Kambi, N. A., Mohanta, S., Raz, A., Haun, A. M., & Saalmann, Y. B. (2021). Consciousness depends on integration between parietal cortex, striatum, and thalamus. *Cell Systems*, *12*(4), 363-373.e11. https://doi.org/10.1016/j.cels.2021.02.003

Alexander, G. E., & Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. *Trends in Neurosciences*, *13*(7), 266–271. https://doi.org/10.1016/0166-2236(90)90107-L

Allen, M., & Friston, K. J. (2018). From cognitivism to autopoiesis: Towards a computational framework for the embodied mind. *Synthese*, *195*(6), 2459–2482. https://doi.org/10.1007/s11229-016-1288-5

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.1

Allen, M., Varga, S., & Heck, D. H. (2022). Respiratory rhythms of the predictive mind. *Psychological Review*. https://doi.org/10.1037/rev0000391

Almeida, L., Deeb, W., Spears, C., Opri, E., Molina, R., Martinez-Ramirez, D., Gunduz, A., Hess, C. W., & Okun, M. S. (2017). Current Practice and the Future of Deep Brain Stimulation Therapy in Parkinson's Disease. *Seminars in Neurology*, *37*(2), 205–214. https://doi.org/10.1055/s-0037-1601893

Andrews, M. (2021). The math is not the territory: Navigating the free energy principle. *Biology & Philosophy*, *36*(3), 30. https://doi.org/10.1007/s10539-021-09807-0

Anggraini, D., Glasauer, S., & Wunderlich, K. (2018). Neural signatures of reinforcement learning correlate with strategy adoption during spatial navigation. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-28241-z

Arsalidou, M., Duerden, E. G., & Taylor, M. J. (2013). The centre of the brain: Topographical model of motor, cognitive, affective, and somatosensory

functions of the basal ganglia. *Human Brain Mapping*, *34*(11), 3031–3054. https://doi.org/10.1002/hbm.22124

Ashby, A., William Ross (1903-1972). (1966). *Design for a Brain: The Origin of Adaptive Behaviour*. Chapman & Hall.

Azzalini, D., Rebollo, I., & Tallon-Baudry, C. (2019). Visceral Signals Shape Brain Dynamics and Cognition. *Trends in Cognitive Sciences*, *23*(6), 488–509. https://doi.org/10.1016/j.tics.2019.03.007

Badcock, P. B., Friston, K. J., & Ramstead, M. J. (2019). The hierarchically mechanistic mind: A free-energy formulation of the human psyche. *Physics of Life Reviews*, *31*, 104-121. https://doi.org/10.1016/j.plrev.2018.10.002

Baldermann, J. C., Schüller, T., Kohl, S., Voon, V., Li, N., Hollunder, B., Figee, M., Haber, S. N., Sheth, S. A., Mosley, P. E., Huys, D., Johnson, K. A., Butson, C., Ackermans, L., Bouwens van der Vlis, T., Leentjens, A. F. G., Barbe, M., Visser-Vandewalle, V., Kuhn, J., & Horn, A. (2021). Connectomic Deep Brain Stimulation for Obsessive-Compulsive Disorder. *Biological Psychiatry*, *90*(10), 678–688. https://doi.org/10.1016/j.biopsych.2021.07.010

Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or Surprise? *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00907

Bell, A. H., Summerfield, C., Morin, E. L., Malecek, N. J., & Ungerleider, L. G. (2016). Encoding of Stimulus Probability in Macaque Inferior Temporal Cortex. *Current Biology*, *26*(17), 2280–2290. https://doi.org/10.1016/j.cub.2016.07.007

Bello-Morales, R., & Delgado-García, J. M. (2015). The social neuroscience and the theory of integrative levels. *Frontiers in Integrative Neuroscience*, *9*. https://doi.org/10.3389/fnint.2015.00054

Bennett, M. R., & Hacker, P. M. S. (2022). *Philosophical Foundations of Neuroscience*. John Wiley & Sons.

Von Bertalanffy, L. (1950). An outline of general system theory. *British Journal for the Philosophy of Science*, *I*(2), 134–165. https://doi.org/10.1093/bjps/I.2.134

Bestmann, S., Harrison, L. M., Blankenburg, F., Mars, R. B., Haggard, P., Friston, K. J., & Rothwell, J. C. (2008). Influence of Uncertainty and Surprise on Human Corticospinal Excitability during Preparation for Action. *Current Biology*, *18*(10), 775–780. https://doi.org/10.1016/j.cub.2008.04.051

Bisenius, S., Trapp, S., Neumann, J., & Schroeter, M. L. (2015). Identifying neural correlates of visual consciousness with ALE meta-analyses. *NeuroImage*, *122*, 177–187. https://doi.org/10.1016/j.neuroimage.2015.07.070

Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology*, *76*, 198–211. https://doi.org/10.1016/j.jmp.2015.11.003

Bole, T. J. (1987). Contradiction in Hegel's Science of Logic. *The Review of Metaphysics*, *40*, 515–534.

B Bolis, D., & Schilbach, L. (2020). 'I interact therefore I am': The self as a historical product of dialectical attunement. *Topoi*, *39*(3), 521-534. https://doi.org/10.1007/s11245-018-9574-0

Boonstra, E. A., van Schouwenburg, M. R., Seth, A. K., Bauer, M., Zantvoord, J. B., Kemper, E. M., Lansink, C. S., & Slagter, H. A. (2020). Conscious perception and the modulatory role of dopamine: No effect of the dopamine D2 agonist cabergoline on visual masking, the attentional blink, and probabilistic discrimination. *Psychopharmacology*, *237*(9), 2855–2872. https://doi.org/10.1007/s00213-020-05579-9

Bordignon, M. (2017). Hegel: A Dialetheist? Truth and Contradiction in Hegel's Logic. *Hegel Bulletin*, *40*(02), 198–214. https://doi.org/10.1017/hgl.2017.15

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brainard, M. S., & Doupe, A. J. (2000). Auditory feedback in learning and maintenance of vocal behaviour. *Nature Reviews Neuroscience*, *1*(1), 31-40. https://doi.org/10.1038/35036205

Breitmeyer, B. (2007). Visual masking: Past accomplishments, present status, future developments. *Advances in Cognitive Psychology*, *3*(1–2), 9–20. https://doi.org/10.2478/v10053-008-0010-7

Brocker, D. T., & Grill, W. M. (2013). Principles of electrical stimulation of neural tissue. In *Handbook of Clinical Neurology* (Vol. 116, pp. 3–18). Elsevier. https://doi.org/10.1016/B978-0-444-53497-2.00001-2

Brown, H., Adams, R. A., Parees, I., Edwards, M., & Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing*, *14*(4), 411–427. https://doi.org/10.1007/s10339-013-0571-3

Brown, L. L., Schneider, J. S., & Lidsky, T. I. (1997). Sensory and cognitive functions of the basal ganglia. *Current Opinion in Neurobiology*, *7*(2), 157–163. https://doi.org/10.1016/S0959-4388(97)80003-7

Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience*, *8*(August), 1–14. https://doi.org/10.3389/fnhum.2014.00599

Bruineberg, J., Rietveld, E., & Kiverstein, J. (2016). The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. *Synthese*. https://doi.org/10.1007/s11229-016-1239-1

Buonomano, D. V., & Merzenich, M. M. (1998). CORTICAL PLASTICITY: From Synapses to Maps. *Annual Review of Neuroscience*, *21*(1), 149–186. https://doi.org/10.1146/annurev.neuro.21.1.149

Buzsáki, G. (2019). *The Brain from Inside Out*. Oxford University Press. https://doi.org/10.1093/oso/9780190905385.001.0001

Cardoso-Leite, P., Mamassian, P., Schütz-Bosbach, S., & Waszak, F. (2010). A New Look at Sensory Attenuation: Action-Effect Anticipation Affects Sensitivity, Not Response Bias. *Psychological Science*, *21*(12), 1740–1745. https://doi.org/10.1177/0956797610389187

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2017). Busting out: Predictive brains, embodied minds, and the puzzle of the evidentiary veil. *Noûs*, *51*(4), 727-753. https://doi.org/10.1111/nous.12140

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

Coenen, V. A., Panksepp, J., Hurwitz, T. A., Urbach, H., & Mädler, B. (2012). Human Medial Forebrain Bundle (MFB) and Anterior Thalamic Radiation (ATR): Imaging of Two Major Subcortical Pathways and the Dynamic Balance of Opposite Affects in Understanding Depression. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *24*(2), 223–236. https://doi.org/10.1176/appi.neuropsych.11080180

Coenen, V. A., Schlaepfer, T. E., Goll, P., Reinacher, P. C., Voderholzer, U., Elst, L. T. van, Urbach, H., & Freyer, T. (2017). The medial forebrain bundle as a target for deep brain stimulation for obsessive-compulsive disorder. *CNS Spectrums*, *22*(3), 282–289. https://doi.org/10.1017/S1092852916000286

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K., & Powers, A. R. (2019). Hallucinations and Strong Priors. *Trends in Cognitive Sciences*, *23*(2), 114–127. https://doi.org/10.1016/j.tics.2018.12.001

Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, *6*(2), 119–126. https://doi.org/10.1038/nn0203-119

Da Costa, L., Parr, T., Sengupta, B., & Friston, K. J. (2021). Neural Dynamics under Active Inference: Plausibility and Efficiency of Information Processing. *Entropy*, *23*(4), Article 4. https://doi.org/10.3390/e23040454

Dandekar, M. P., Fenoy, A. J., Carvalho, A. F., Soares, J. C., & Quevedo, J. (2018). Deep brain stimulation for treatment-resistant depression: an integrative review of preclinical and clinical findings and translational implications. *Molecular psychiatry*, *23*(5), 1094-1112. https://doi.org/10.1038/mp.2018.2

Daunizeau, J., Ouden, H. E. M. den, Pessiglione, M., Kiebel, S. J., Friston, K. J., & Stephan, K. E. (2010). Observing the Observer (II): Deciding When to Decide. *PLOS ONE*, *5*(12), e15555. https://doi.org/10.1371/journal.pone.0015555

Dehaene, S., & Changeux, J.-P. (2011). Experimental and Theoretical Approaches to Conscious Processing. *Neuron*, *70*(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018

Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology*, *5*(10), 2408–2423. https://doi.org/10.1371/journal.pbio.0050260

Dombert, P. L., Kuhns, A., Mengotti, P., Fink, G. R., & Vossel, S. (2016). Functional mechanisms of probabilistic inference in feature- and space-based attentional systems. *NeuroImage*, *142*, 553–564. https://doi.org/10.1016/j.neuroimage.2016.08.010

Edelman, G. M. (2003). Naturalizing consciousness: A theoretical framework. *Proceedings of the National Academy of Sciences*, *100*(9), 5520–5524. https://doi.org/10.1073/pnas.0931349100

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 229.

Feldman, D. E. (2009). Synaptic Mechanisms for Plasticity in Neocortex. *Annual Review of Neuroscience*, *32*(1), 33–55. https://doi.org/10.1146/annurev.neuro.051508.135516

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215. https://doi.org/10.3389/fnhum.2010.00215

Figee, M., Luigjes, J., Smolders, R., Valencia-Alfonso, C. E., Van Wingen, G., De Kwaasteniet, B., ... & Denys, D. (2013). Deep brain stimulation restores frontostriatal network activity in obsessive-compulsive disorder. *Nature neuroscience*, *16*(4), 386-387. https://doi.org/10.1038/nn.3344

Fisher, M. (2009). *Capitalist Realism: Is There No Alternative?* John Hunt Publishing.

FitzGerald, T. H. B., Dolan, R. J., & Friston, K. (2015). Dopamine, reward learning, and active inference. *Frontiers in Computational Neuroscience*, *9*. https://doi.org/10.3389/fncom.2015.00136

Frässle, S., Aponte, E. A., Bollmann, S., Brodersen, K. H., Do, C. T., Harrison, O. K., Harrison, S. J., Heinzle, J., Iglesias, S., Kasper, L., Lomakina, E. I., Mathys, C., Müller-Schrader, M., Pereira, I., Petzschner, F. H., Raman, S., Schöbi, D., Toussaint, B., Weber, L. A., … Stephan, K. E. (2021). TAPAS: An Open-Source Software Package for Translational Neuromodeling and Computational Psychiatry. *Frontiers in Psychiatry*, *12*. https://www.frontiersin.org/article/10.3389/fpsyt.2021.680811

Fridgeirsson, E. A., Bais, M. N., Eijsker, N., Thomas, R. M., Smit, D. J. A., Bergfeld, I. O., … & Denys, D. (2021). The ventral striatum harbours patient specific intracranial neural signatures of obsessions and compulsions. *medRxiv*. https://doi.org/10.1101/2021.04.04.21254830

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. (2012). A free energy principle for biological systems. *Entropy*, *14*(11), 2100–2121. https://doi.org/10.3390/e14112100

Friston, K. (2018a). Am I self-conscious? *Frontiers in Psychology*, *9*(April), 579–579. https://doi.org/10.3389/FPSYG.2018.00579

Friston, K. (2018b). Does predictive coding have a future? *Nature Neuroscience*, *21*(8), 1019–1021. https://doi.org/10.1038/s41593-018-0200-7

Friston, K. (2019). A free energy principle for a particular physics. *ArXiv:1906.10184 [q-Bio]*. http://arxiv.org/abs/1906.10184

Friston, K., & Buzsáki, G. (2016). The Functional Anatomy of Time: What and When in the Brain. *Trends in Cognitive Sciences*, *20*(7), 500–511. https://doi.org/10.1016/j.tics.2016.05.001

Friston, K. (2013). Life as we know it. *Journal of the Royal Society Interface*, *10*(86), 20130475. https://doi.org/10.1038/449001a

Friston, K. J. (2013). Active inference and free energy. *Behavioral and Brain Sciences*, *36*(3), 212–213. https://doi.org/10.1017/S0140525X12002142

Friston, K. J., Da Costa, L., Hafner, D., Hesp, C., & Parr, T. (2021). Sophisticated Inference. *Neural Computation*, *33*(3), 713–763. https://doi.org/10.1162/neco_a_01351

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., Dolan, R. J., Moran, R., Stephan, K. E., & Bestmann, S. (2012). Dopamine,

affordance and active inference. *PLoS Computational Biology*, *8*(1). https://doi.org/10.1371/journal.pcbi.1002327

Friston, K. J., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: The brain as a phantastic organ. *The Lancet Psychiatry*, *1*(2), 148–158. https://doi.org/10.1016/S2215-0366(14)70275-5

Friston, K. J., Wiese, W., & Hobson, J. A. (2020). Sentience and the Origins of Consciousness: From Cartesian Duality to Markovian Monism. *Entropy*, *22*(5), 516. https://doi.org/10.3390/e22050516

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2014). The anatomy of choice: Dopamine and decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130481–20130481. https://doi.org/10.1098/rstb.2013.0481

Friston, K., Thornton, C., & Clark, A. (2012). Free-energy minimization and the dark-room problem. *Frontiers in psychology*, 130. https://doi.org/10.3389/fpsyg.2012.00130

Gershman, S. J. (2019). What does the free energy principle tell us about the brain? *Neurons, Behavior, Data analysis, and Theory*, *2*(3), 1-10.

Graat, I., Liebrand, L., van de Munckhof, P., Schuurman, R., Bergfeld, I., Mocking, R., & Denys, D. (2021). Effect of brain target and stimulation paradigm on response to deep brain stimulation in obsessive-compulsive disorder. *Brain Stimulation*, *14*(6), 1727–1728. https://doi.org/10.1016/j.brs.2021.10.461

Haber, S. N., & McFarland, N. R. (1999). The Concept of the Ventral Striatum in Nonhuman Primates. *Annals of the New York Academy of Sciences*, *877*(1), 33–48. https://doi.org/10.1111/j.1749-6632.1999.tb09259.x

Hegel, G. W. F. (2007). *Georg Wilhelm Friedrich Hegel: Lectures on the Philosophy of Spirit 1827-8*. Oxford University Press.

Hegel, G. W. F., & Miller, A. V. (2004). *Hegel's Philosophy of Nature: Being Part Two of the Encyclopedia of the Philosophical Sciences (1830), Translated from Nicolin and Pöggeler's Edition (1959), and from the Zusätze in Michelet's Text (1847)*. Clarendon Press.

Helmholtz, H. von. (1995). *Science and Culture: Popular and Philosophical Essays* (D. Cahan, Ed.). University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/S/bo3616257.html

Hobson, J. A., & Friston, K. J. (2014). Consciousness, dreams, and inference: The Cartesian theatre revisited. *Journal of Consciousness Studies*, *21*(1-2), 6-32.

Hobson, J. A., & Friston, K. J. (2016). A response to our theatre critics. *Journal of Consciousness Studies*, *23*(3-4), 245-254.

Hobson, J. A., Hong, C. C.-H., & Friston, K. J. (2014). Virtual reality and consciousness inference in dreaming. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/article/10.3389/fpsyg.2014.01133

Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199682737.001.0001

Hohwy, J. (2016). The self-evidencing brain. *Nous*, *50*(2), 259–285. https://doi.org/10.1111/nous.12062

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, *1*(II). https://doi.org/10.33735/phimisci.2020.II.64

Hommel, B. (1996). The cognitive representation of action: Automatic integration of perceived action effects. *Psychological Research*, *59*(3), 176–186. https://doi.org/10.1007/BF00425832

Hommel, B. (2019). Theory of Event Coding (TEC) V2. 0: Representing and controlling perception and action. *Attention, Perception, & Psychophysics*, *81*(7), 2139-2154. https://doi.org/10.3758/s13414-019-01779-4

Hommel, B., & Elsner, B. (2009). Acquisition, representation, and control of action. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (Eds.), *Oxford Handbook of Human Action* (pp. 368–397). Oxford University Press.

Kant, I. (2000). *Critique of the Power of Judgment*. Cambridge University Press.

Kennedy, B. L., Rawding, J., Most, S. B., & Hoffman, J. E. (2014). Emotion-induced blindness reflects competition at early and late processing stages: An ERP study. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(4), 1485–1498. https://doi.org/10.3758/s13415-014-0303-x

Kiverstein, J., Rietveld, E., Slagter, H. A., & Denys, D. (2019). Obsessive Compulsive Disorder: A Pathology of Self-Confidence? *Trends in Cognitive Sciences*, *23*(5), 369–372. https://doi.org/10.1016/j.tics.2019.02.005

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, *75*(2), 265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Korka, B., Schröger, E., & Widmann, A. (2019). Action Intention-based and Stimulus Regularity-based Predictions: Same or Different? *Journal of Cognitive Neuroscience*, *31*(12), 1917–1932. https://doi.org/10.1162/jocn_a_01456

Korka, B., Widmann, A., Waszak, F., Darriba, Á., & Schröger, E. (2022). The auditory brain in action: Intention determines predictive processing

in the auditory system—A review of current paradigms and findings. *Psychonomic Bulletin & Review*, *29*(2), 321–342. https://doi.org/10.3758/s13423-021-01992-z

Kuhns, A. B., Dombert, P. L., Mengotti, P., Fink, G. R., & Vossel, S. (2017). Spatial Attention, Motor Intention, and Bayesian Cue Predictability in the Human Brain. *Journal of Neuroscience*, *37*(21), 5334–5344. https://doi.org/10.1523/JNEUROSCI.3255-16.2017

Leuthold, H. (2011). The Simon effect in cognitive electrophysiology: A short review. *Acta Psychologica*, *136*(2), 203–211. https://doi.org/10.1016/j.actpsy.2010.08.001

Lewontin, R., & Levins, R. (2007). *Biology Under the Influence: Dialectical Essays on the Coevolution of Nature and Society*. NYU Press.

Liebrand, L. C., Caan, M. W. A., Schuurman, P. R., Munckhof, P. van den, Figee, M., Denys, D., & Wingen, G. A. van. (2019). Individual white matter bundle trajectories are associated with deep brain stimulation response in obsessive-compulsive disorder. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, *12*(2), 353–360. https://doi.org/10.1016/j.brs.2018.11.014

Locke, J. (1796). *An Essay Concerning Human Understanding*. T. Longman.

López-Astorga, M. (2016). The principle of explosion: Aristotle versus the current syntactic theories. *ΣΧΟΛΗ. Философское антиковедение и классическая традиция*, *10*(1), 40-49.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Smíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E.-J. (2019). **JASP**: Graphical Statistical Software for Common Statistical Designs. *Journal of Statistical Software*, *88*(2). https://doi.org/10.18637/jss.v088.i02

Malabou, C. (2005). *The Future of Hegel: Plasticity, Temporality, and Dialectic*. Psychology Press.

Malabou, C. (2008). *What Should We Do with Our Brain?* Fordham University Press.

Marques, V. (2016). Positing the Presuppositions—Dialectical Biology and the Minimal Structure of Life. In A. Hamza & F. Ruda (Eds.), *Slavoj Žižek and Dialectical Materialism* (pp. 113–132). Palgrave Macmillan US. https://doi.org/10.1057/9781137538611_9

Marques, V., & Brito, C. (2014). THE RISE AND FALL OF THE MACHINE METAPHOR: ORGANIZATIONAL SIMILARITIES AND DIFFERENCES BETWEEN MACHINES AND LIVING BEINGS. *Verifiche XLIII*, *1*(3), 77–111.

Marx, K., & Engels, F. (2002). *The Communist Manifesto.* Penguin Publishing Group.

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian Foundation for Individual Learning Under Uncertainty. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00039

Mathys, C., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00825

Maturana, H. R., & Varela, F. J. (2012). *Autopoiesis and Cognition: The Realization of the Living*. Springer Science & Business Media.

McIntyre, C. C., Grill, W. M., Sherman, D. L., & Thakor, N. V. (2004). Cellular Effects of Deep Brain Stimulation: Model-Based Analysis of Activation and Inhibition. *Journal of Neurophysiology*, *91*(4), 1457–1469. https://doi.org/10.1152/jn.00989.2003

Michelini, F. (2012). Hegel's notion of natural purpose. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 133–139. https://doi.org/10.1016/j.shpsc.2011.05.012

Michelini, F., Wunsch, M., & Stederoth, D. (2018). Philosophy of nature and organism's autonomy: On Hegel, Plessner and Jonas' theories of living beings. *History and Philosophy of the Life Sciences*, *40*(3), 56. https://doi.org/10.1007/s40656-018-0212-3

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144. https://doi.org/10.1016/S1364-6613(03)00029-9

Most, S. B., Chun, M. M., Widders, D. M., & Zald, D. H. (2005). Attentional rubbernecking: Cognitive control and personality in emotion-induced blindness. *Psychonomic Bulletin & Review*, *12*(4), 654–661. https://doi.org/10.3758/BF03196754

Nittono, H. (2004, August). The action-perception paradigm: A new perspective in cognitive neuroscience. In *International Congress Series* (Vol. 1270, pp. 26-31). Elsevier. https://doi.org/10.1016/j.ics.2004.04.093

Nittono, H., Hamada, A., & Hori, T. (2003). Brain Potentials after Clicking a Mouse: A New Psychophysiological Approach to Human-Computer Interaction. *Human Factors*, *45*(4), 591–600. https://doi.org/10.1518/hfes.45.4.591.27087

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, *42*(1), 188–204. https://doi.org/10.3758/BRM.42.1.188

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Ostry, D. J., & Gribble, P. L. (2016). Sensory Plasticity in Human Motor Learning. *Trends in Neurosciences*, *39*(2), 114–123. https://doi.org/10.1016/j.tins.2015.12.006

Parr, T., & Friston, K. J. (2019). Generalised free energy and active inference. *Biological cybernetics*, *113*(5), 495-513. https://doi.org/10.1007/s00422-019-00805-w

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior.* MIT Press.

Pezzulo, G., Rigoli, F., & Friston, K. J. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17–35. https://doi.org/10.1016/j.pneurobio.2015.09.001

Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: Knowns and unknowns. *Nature Neuroscience*, *16*(9), 1170–1178. https://doi.org/10.1038/nn.3495

Press, C., Kok, P., & Yon, D. (2020). The Perceptual Prediction Paradox. *Trends in Cognitive Sciences*, *24*(1), 13–24. https://doi.org/10.1016/j.tics.2019.11.003

Prinz, W. (1997). Perception and Action Planning. *European Journal of Cognitive Psychology*, *9*(2), 129–154. https://doi.org/10.1080/713752551

Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2018). Answering Schrödinger's question: A free-energy formulation. *Physics of Life Reviews*, *24*, 1–16. https://doi.org/10.1016/j.plrev.2017.09.001

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860. https://doi.org/10.1037/0096-1523.18.3.849

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience*, *89*(4), 1009–1023. https://doi.org/10.1016/S0306-4522(98)00319-4

Redgrave, P., Vautrelle, N., & Reynolds, J. N. J. (2011). Functional properties of the basal ganglia's re-entrant loop architecture: selection and reinforcement. *Neuroscience*, *198*, 138-151. https://doi.org/10.1016/j.neuroscience.2011.07.060

Reznik, D., Henkin, Y., Schadel, N., & Mukamel, R. (2014). Lateralized enhancement of auditory cortex activity and increased sensitivity to self-generated sounds. *Nature Communications*, *5*(1), 4059. https://doi.org/10.1038/ncomms5059

Reznik, D., & Mukamel, R. (2019). Motor output, neural states and auditory perception. *Neuroscience & Biobehavioral Reviews*, *96*, 116–126. https://doi.org/10.1016/j.neubiorev.2018.10.021

Sajid, N., Ball, P. J., Parr, T., & Friston, K. J. (2021). Active Inference: Demystified and Compared. *Neural Computation*, *33*(3), 674–712. https://doi.org/10.1162/neco_a_01357

Schafer, E. W., & Marcus, M. M. (1973). Self-stimulation alters human sensory brain responses. *Science*, *181*(4095), 175-177.

Schrodinger, E. (2012). *What is Life? With Mind and Matter and Autobiographical Sketches*. Cambridge University Press.

Schwarz, K. A., Pfister, R., Kluge, M., Weller, L., & Kunde, W. (2018). Do we see it or not? Sensory attenuation in the visual domain. *Journal of Experimental Psychology: General*, *147*(3), 418–430. https://doi.org/10.1037/xge0000353

Seger, C. A. (2013). The visual corticostriatal loop through the tail of the caudate: Circuitry and function. *Frontiers in Systems Neuroscience*, *7*(December), 1–15. https://doi.org/10.3389/fnsys.2013.00104

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*(10), Article 10. https://doi.org/10.1038/nn1549

Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, *5*(2), 97–118. https://doi.org/10.1080/17588928.2013.877880

Seth, A. K. (2015). The Cybernetic Bayesian BrainThe Cybernetic Bayesian Brain: From Interoceptive Inference to Sensorimotor Contingencies: From Interoceptive Inference to Sensorimotor Contingencies. *Open MIND*. https://doi.org/10.15502/9783958570108

Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*, 1-14. https://doi.org/10.1038/s41583-022-00587-4

Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160007. https://doi.org/10.1098/rstb.2016.0007

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1997). The attentional blink. *Trends in Cognitive Sciences*, *1*(8), 291–296. https://doi.org/10.1016/S1364-6613(97)01094-2

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*(3), 300–304. https://doi.org/10.1037/h0020586

Slagter, H. A., Davidson, R. J., & Tomer, R. (2010). Eye-blink rate predicts individual differences in pseudoneglect. *Neuropsychologia*, *48*(5), 1265–1268. https://doi.org/10.1016/j.neuropsychologia.2009.12.027

Slagter, H. A., Mazaheri, A., Reteig, L. C., Smolders, R., Figee, M., Mantione, M., Schuurman, P. R., & Denys, D. (2017). Contributions of the Ventral Striatum to Conscious Perception: An Intracranial EEG Study of the Attentional Blink. *The Journal of Neuroscience*, *37*(5), 1081–1089. https://doi.org/10.1523/JNEUROSCI.2282-16.2016

Slagter, H. A., Tomer, R., Christian, B. T., Fox, A. S., Colzato, L. S., King, C. R., Murali, D., & Davidson, R. J. (2012). PET Evidence for a Role for Striatal Dopamine in the Attentional Blink: Functional Implications. *Journal of Cognitive Neuroscience*, *24*(9), 1932–1940. https://doi.org/10.1162/jocn_a_00255

Smith, Y., Raju, D. V., Pare, J.-F., & Sidibe, M. (2004). The thalamostriatal system: A highly specific network of the basal ganglia circuitry. *Trends in Neurosciences*, *27*(9), 520–527. https://doi.org/10.1016/j.tins.2004.07.004

Smolders, R., Mazaheri, A., van Wingen, G., Figee, M., de Koning, P. P., & Denys, D. (2013). Deep Brain Stimulation Targeted at the Nucleus Accumbens Decreases the Potential for Pathologic Network Communication. *Biological Psychiatry*, *74*(10), e27–e28. https://doi.org/10.1016/j.biopsych.2013.03.012

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017. https://doi.org/10.1016/j.neuroimage.2009.03.025

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.

Swanson, L. R. (2016). The predictive processing paradigm has roots in Kant. *Frontiers in Systems Neuroscience*, *10*, 79. https://doi.org/10.3389/fnsys.2017.00098

Sweatt, J. D. (2016). Neural plasticity and behavior – sixty years of conceptual advances. *Journal of Neurochemistry*, *139*, 179–199. https://doi.org/10.1111/jnc.13580

Synofzik, M., Vosgerau, G., & Voss, M. (2013). The experience of agency: An interplay between prediction and postdiction. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00127

Thaler, L., Schütz, A. C., Goodale, M. A., & Gegenfurtner, K. R. (2013). What is the best fixation target? The effect of target shape on stability of fixational

eye movements. *Vision Research*, *76*, 31–42. https://doi.org/10.1016/j.visres.2012.10.012

Thompson, E., & Varela, F. J. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences*, *5*(10), 418–425. https://doi.org/10.1016/S1364-6613(00)01750-2

van den Heuvel, O. A., van Wingen, G., Soriano-Mas, C., Alonso, P., Chamberlain, S. R., Nakamae, T., Denys, D., Goudriaan, A. E., & Veltman, D. J. (2016). Brain circuitry of compulsivity. *European Neuropsychopharmacology*, *26*(5), 810–827. https://doi.org/10.1016/j.euroneuro.2015.12.005

van Opstal, F., Van Laeken, N., Verguts, T., Van Dijck, J. P., De Vos, F., Goethals, I., & Fias, W. (2014). Supp: Correlation between individual differences in striatal dopamine and in visual consciousness. *Current Biology*, *24*(7). https://doi.org/10.1016/j.cub.2014.02.001

van Wingen, G., Bergfeld, I., de Koning, P., Graat, I., Luigjes, J., Mocking, R., … & Denys, D. (2022). Comment to: Deep brain stimulation for refractory obsessive-compulsive disorder (OCD): emerging or established therapy? *Molecular Psychiatry*, *27*(3), 1276-1277. https://doi.org/10.1038/s41380-021-01411-8

Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition. *Brain and Cognition*, *34*(1), 72-87.

Vasser, M., Vuillaume, L., Cleeremans, A., & Aru, J. (2019). Waving goodbye to contrast: Self-generated hand movements attenuate visual sensitivity. *Neuroscience of Consciousness*, *2019*(1). https://doi.org/10.1093/nc/niy013

Vossel, S., Bauer, M., Mathys, C., Adams, R. A., Dolan, R. J., Stephan, K. E., & Friston, K. J. (2014). Cholinergic Stimulation Enhances Bayesian Belief Updating in the Deployment of Spatial Attention. *Journal of Neuroscience*, *34*(47), 15735–15742. https://doi.org/10.1523/JNEUROSCI.0091-14.2014

Vossel, S., Mathys, C., Daunizeau, J., Bauer, M., Driver, J., Friston, K. J., & Stephan, K. E. (2014). Spatial Attention, Precision, and Bayesian Inference: A Study of Saccadic Response Speed. *Cerebral Cortex*, *24*(6), 1436–1450. https://doi.org/10.1093/cercor/bhs418

Vossel, S., Mathys, C., Stephan, K. E., & Friston, K. J. (2015). Cortical Coupling Reflects Bayesian Belief Updating in the Deployment of Spatial Attention. *The Journal of Neuroscience*, *35*(33), 11532–11542. https://doi.org/10.1523/JNEUROSCI.1382-15.2015

Weber, A., & Varela, F. J. V. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, *1*, 97–125. https://doi.org/10.1023/A:1020368120174

Weiss, C., Herwig, A., & Schütz-Bosbach, S. (2011). The self in action effects: Selective attenuation of self-generated sounds. *Cognition*, *121*(2), 207–218. https://doi.org/10.1016/j.cognition.2011.06.011

Weiss, C., & Schütz-Bosbach, S. (2012). Vicarious action preparation does not result in sensory attenuation of auditory action effects. *Consciousness and Cognition*, *21*(4), 1654–1661. https://doi.org/10.1016/j.concog.2012.08.010

Wen, W. (2019). Does delay in feedback diminish sense of agency? A review. *Consciousness and Cognition*, *73*, 102759. https://doi.org/10.1016/j.concog.2019.05.007

Whiting, A. C., Oh, M. Y., & Whiting, D. M. (2018). Deep brain stimulation for appetite disorders: A review. *Neurosurgical Focus*, *45*(2), E9. https://doi.org/10.3171/2018.4.FOCUS18141

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data – a practical approach. *Journal of Neuroscience Methods*, *250*, 34–46. https://doi.org/10.1016/j.jneumeth.2014.08.002

Wiese, W., & Metzinger, T. (2017). Vanilla predictive processing for philosophers: A primer on predictive processing. *Philosophy and Predictive Processing*, 1–18. https://doi.org/10.15502/9783958573024

Wittenberg, G. F. (2010). Experience, cortical remapping, and recovery in brain disease. *Neurobiology of Disease*, *37*(2), 252–258. https://doi.org/10.1016/j.nbd.2009.09.007

Wu, H., Hariz, M., Visser-Vandewalle, V., Zrinzo, L., Coenen, V. A., Sheth, S. A., ... & Nuttin, B. (2021). Deep brain stimulation for refractory obsessive-compulsive disorder (OCD): emerging or established therapy?. *Molecular psychiatry*, *26*(1), 60-65. https://doi.org/10.1038/s41380-020-00933-x

Žižek, S. (1989). *The Sublime Object of Ideology*. Verso Books.

Žižek, S. (1993). *Tarrying with the Negative: Kant, Hegel, and the Critique of Ideology*. Duke University Press.

Žižek, S. (2009). *The Parallax View*. MIT Press.

Žižek, S. (2012). *Less Than Nothing: Hegel and the Shadow of Dialectical Materialism*. Verso Books.

Žižek, S. (2016). *Disparities*. Bloomsbury Publishing.