



MONASH University

**Allostasis and uncertainty:
An active inference perspective**

Andrew William Corcoran

A thesis submitted for the degree of *Doctor of Philosophy* at
Monash University in 2021

Philosophy Department
School of Philosophical, Historical and International Studies
Faculty of Arts, Monash University

Copyright notice

© Andrew William Corcoran (2021). Except as provided in the Copyright Act 1968, this thesis may not be reproduced in any form without the written permission of the author.

Under the Copyright Act 1968, this thesis must be used only under the normal conditions of scholarly fair dealing. In particular, no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Abstract

Biological systems embody a morass of precariously balanced parts and processes, yet remain remarkably resilient to perturbation. This characteristic ability to stave off disorder is often realised through actions that preserve the system’s functional organisation. The aim of this thesis is to provide an account of such regulatory activity from the perspective of active inference.

Active inference presents a formal framework that conceptualises the emergence of adaptive dynamics in terms of uncertainty reduction. Put simply, this perspective encourages the view that all expressions of adaptive agency entail the resolution of uncertainty. This perspective, which is derived under the free energy principle in accordance with the normative prescriptions of Bayesian inference, has far-reaching implications for the way biological self-organisation and activity are understood.

This thesis grapples with the implications of active inference for biological regulation in cognitive agents, focusing in particular on a predictive mode of regulation: *allostasis*. It begins by considering the free energy principle’s foundational concern with homeostasis, and the way active inference has informed theoretical accounts of interoception and allostatic control. This analysis is then extended to consider how different modes of biological regulation might be formalised under active inference, leading to a proposal for differentiating cognitive from non-cognitive agents.

Next, the relation between cognition and physiological regulation is re-examined from the perspective of embodiment. I challenge the notion that active inference solves intractable debates about the embodiment of mind. I then propose a novel hypothesis that grounds the emergence of cognition in biological rhythms such as the heartbeat.

The second half of the thesis presents three empirical studies investigating the role of covert attentional processes in uncertainty reduction. A binocular rivalry experiment provides evidence of brain-heart integration whereby cardiac activity is progressively modulated in response to increasing perceptual uncertainty. Evidence of cardiac coupling with sensorimotor activity is also reported in the context of mind-wandering. Finally, the covert adaptation of neural oscillations is addressed in a study of degraded speech perception.

In sum, this thesis argues for a unified active inference account which conceives of biological regulation and cognition as integrated modes of uncertainty reduction.

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes two original papers published in peer reviewed journals and one original chapter published in an edited collection. The core theme of the thesis is the biological regulation of uncertainty as understood under the active inference framework. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, Andrew Corcoran, working within the School of Philosophical, Historical and International Studies under the primary supervision of Professor Jakob Hohwy, and the secondary supervision of Professor Vaughan Macefield of the Baker Heart and Diabetes Institute.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research. In the case of the published works featured in Chapter 2, Chapter 3, and Chapter 5, the contribution of myself and my co-authors is summarised in Table 1. For the two unpublished experiments reported in Chapter 6 and Chapter 7, I was responsible for leading data analysis and manuscript writing; I also collected data for Chapter 7. Further assistance with data collection was provided by various research assistants under the supervision of myself and Dr. Thomas Andrillon. Dr. Andrillon also contributed methodological expertise for the stimulus reconstruction analysis reported in Chapter 7, while Mr. Ricardo Perera assisted with the construction of sentence stimuli. The additionally-listed co-authors were involved in project conception.

I have renumbered pages (but not sections) of published papers in order to generate a consistent presentation within the thesis.

Student name: Andrew W. Corcoran

Student signature:

Date: 01/06/2021

I hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not

the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor name: Jakob Hohwy

Main Supervisor signature:

Date: 01/06/2021

Chapter	Publication Title	Status	Nature and % of student contribution	Co-author name(s) Nature and % of Co-author's contribution	Co-author(s), Monash student (Y/N)
2	Allostasis, interoception, and the free energy principle: Feeling our way forward	Published	90% – Concept, original draft, revision	Jakob Hohwy – 10% Concept, revision	No
3	From allostatic agents to counterfactual cognisers: Active inference, biological regulation, and the origins of cognition	Published	85% – Concept, original draft, revision	Jakob Hohwy – 10% Concept, revision Giovanni Pezzulo – 5% Concept, revision	No No
5	Be still my heart: Cardiac regulation as a mode of uncertainty reduction	Published	85% – Concept, data curation & analysis, original draft, revision	Jakob Hohwy – 10% Concept, design, revision Vaughan Macefield – 5% Design, analysis, revision	No No

TABLE 1: Contribution to published works included within thesis.

Publications during enrolment

- **Corcoran, A. W.**, Macefield, V. G., & Hohwy, J. (2021). Be still my heart: Cardiac regulation as a mode of uncertainty reduction. *Psychonomic Bulletin & Review*, 28(4), 1211–1223.
- Cross, Z. R., Santamaria, A., **Corcoran, A. W.**, Chatburn, A., Alday, P. M., Coussens, S., & Kohler, M. J. (2020). Individual alpha frequency modulates sleep-related emotional memory consolidation. *Neuropsychologia*, 148, 107660.
- Engel, M. M., van Denderen, K., Bakker, A.-R., **Corcoran, A. W.**, Keiser, A., & Dijkerman, H. C. (2020). Anorexia and the size-weight illusion: No evidence of impaired visual-haptic object integration. *PLoS ONE*, 15(8), e0237421.
- **Corcoran, A. W.**, Pezzulo, G., & Hohwy, J. (2020). From allostatic agents to counterfactual cognisers: Active inference, biological regulation, and the origins of cognition. *Biology & Philosophy*, 35, 32.
- **Corcoran, A. W.** (2019). Cephalopod molluscs, causal models, and curious minds. *Animal Sentience*, 4(26), 13.
- Parr, T., **Corcoran, A. W.**, Friston, K. J., & Hohwy, J. (2019). Perceptual awareness and active inference. *Neuroscience of Consciousness*, 5(1), niz012.
- **Corcoran, A. W.**, Pezzulo, G., & Hohwy, J. (2018). Commentary: Respiration-entrained brain rhythms are global but often overlooked. *Frontiers in Systems Neuroscience*, 12, 25.
- **Corcoran, A. W.** & Hohwy, J. (2018). Allostasis, interoception, and the free energy principle: Feeling our way forward. In M. Tsakiris & H. De Preester (Eds.), *The interoceptive mind: From homeostasis to awareness* (pp. 272–292). Oxford: Oxford University Press
- **Corcoran, A. W.**, Groot, C., Bruno, A., Johnston, A., & Cropper, S. J. (2018). Individual differences in first- and second-order temporal judgment. *PLoS ONE*, 13(2), e0191422.
- **Corcoran, A. W.**, Alday, P. M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2018). Toward a reliable, automated method of individual alpha frequency (IAF) quantification. *Psychophysiology*, 55(7), e13064.

Acknowledgements

I expected the doctorate to be difficult; I didn't expect it to be *this* difficult. That this thesis exists is a testament to a great many people who have supported, encouraged, and cajoled me along the way – the past 4 years and beyond.

First and foremost, I must express my sincere gratitude to my primary supervisor, Prof. Jakob Hohwy. Jakob has been as good a mentor and lab head as anyone could hope for during a doctorate. Not only is he a bone fide leader in his field, he is a remarkably humble, good-humoured, and kind man – one whose door has always been open despite an endless torrent of administrative duties. Above all, I am deeply indebted to Jakob for his honesty, encouragement, and patience – always granting me the freedom to go wherever the ideas led, but knowing when the time was right to set me back on course.

One of the many privileges of being a student of Jakob's is that it opens many doors; during my doctorate I've been fortunate enough to be introduced to a great many luminaries, some of whom I've had the great pleasure of collaborating with. Amongst these, I am especially grateful to Prof. Vaughan Macefield for lending his considerable expertise in the capacity of associate supervisor – expertise that was invaluable for transforming my nebulous ideas into rigorous psychophysiological research.

As for unofficial mentors, I've had a few. My sincere thanks to Dr. Giovanni Pezzulo and his team at L'Istituto di Scienze e Tecnologie della Cognizione for their generous hospitality during a visit to Rome in January 2018. The many stimulating conversations shared during that trip had a profound impact on my thinking, the traces of which permeate the pages of this document. Thanks also to Dr. Thomas Andrillon not only for his vital contribution to this thesis, but also for his warmth and generosity of spirit through countless video calls.

Further afield, I cannot begin to express my appreciation to Profs. Ina Bornkessel-Schlesewsky and Matthias Schlewsky for their enduring faith and kindness, without which I never would have embarked upon this project. Not only did my time with them – and their protégé Dr. Phillip Alday – introduce me to a strange and beguiling principle that would come to dominate my doctoral research, it equipped me with the (misplaced) confidence and skills I would need to pursue it.

Membership of the Cognition & Philosophy Lab also confers its fair share of informal guides and mentors. I benefitted greatly in the early days of my candidature from the

experience and advice of Drs. Noam Gordon and Julian Matthews, who at the time were nearing the final stages of their doctoral research. I also received invaluable help on numerous occasions from Drs. Mateusz Woźniak, Roger Koenig-Robert, and Bryan Paton, all of whom command a staggering breadth and depth of knowledge.

More broadly, it has been a pleasure and a privilege to belong to such a thriving and intellectually challenging community, and to meet so many brilliant people – far too many to mention. Of our many international visitors, two are worthy of special mention – Drs. Thomas Parr and Dan Williams, men of supreme intellect and humility whose work has deeply influenced my own (far less impressive) endeavours. From the faculty, thanks to Prof. Tim Bayne and Dr. Jenny Windt for their perennially insightful questions and comments (even if I could seldom do them justice). And thanks especially to Dr. Monima Chadha for always taking an interest, and always having a moment.

To those I've been fortunate enough to share an office with, and countless lunches, chats and coffees – thank you all. To the philosophers – Steve, Andy, Iwan, Niccolò, Ricardo – thanks for putting up with my lumpen ignorance, and for gently nudging me in the right direction when you could. To the non-philosophers – Simon, Gus, Jonno – thanks for not being philosophers (and yet still putting up with my lumpen ignorance). And to my virtual labmates, Zach and Louise – thank you both for always being there, if not here, even before the beginning.

But of all the labmates, my deepest appreciation is reserved for three outstanding scientists who also happen to be three of the best people I know. Kelsey – from the beginning you were always the one who set the benchmark; not just a paragon of what any young scientist should aspire to, but someone who knits the lab together in so many ways. Manu – you're one of the most extraordinarily talented people I know. I deeply regret not getting to know you sooner. Manja – well I know I don't have to say it, but I will: your European work ethic really is quite remarkable. At any rate, all three of you were there for me when it really mattered; I'll be eternally grateful for that.

Finally, and above all, thanks to my family – Alison, Alex, and the kittens for harbouring me during those transits through Europe, and always making me feel at home. Sheila, for teaching me more about the relation between body and mind than anyone. My parents, Ruth and John, for asking about the papers and timeline, the last experiment and the next, and always having my best interests at heart. And thank you at last to Winnie and Alisa – for everything.

Perhaps the most valuable result of all education is the ability to make yourself do the thing you have to do, when it ought to be done, whether you like it or not; it is the first lesson that ought to be learned; and however early a man's training begins, it is probably the last lesson that he learns thoroughly.

THOMAS HENRY HUXLEY

Contents

Copyright notice	i
Abstract	ii
Thesis including published works declaration	iii
Publications during enrolment	vi
Acknowledgements	vii
1 Introduction	1
1.1 The free energy principle	2
1.2 Interoceptive inference	2
1.3 Active sensing and covert action	3
1.4 Overview of thesis structure	4
2 Allostasis, interoception, and the free energy principle: Feeling our way forward	7
2.1 Introduction	8
2.2 Discovering “the wisdom of the body”: Homeostasis	9
2.3 Allostasis: The future of homeostatic regulation?	10
2.3.1 Achieving stability through change	11
2.3.2 Allostatic means for homeostatic ends	11
2.3.3 Two modes of sustained viability	13
2.4 Allostasis and interoceptive inference	14
2.4.1 Behavioral allostasis	14
2.4.2 Teleological allostasis	18
2.4.3 Diachronic allostasis	21
2.5 The future of the history of allostasis	24
2.6 References	25
3 From allostatic agents to counterfactual cognisers: Active inference, biological regulation, and the origins of cognition	29
3.1 Introduction	31
3.2 Homeostasis and the free energy principle	32
3.2.1 Life, formalised: thermodynamics, attracting sets, and (un)certainly	32
3.2.2 Surprise and free energy minimisation	33
3.2.3 Existence implies inference: agents as generative, self-evidencing models	35

3.2.4	Active inference: closing the perception–action loop	36
3.3	Beyond homeostasis: allostasis and hierarchical generative models	38
3.3.1	Allostasis under active inference	39
3.3.2	Broadening the inferential horizon: preferences, policies, and plans	41
3.3.3	Interim summary	44
3.4	Biological regulation in an uncertain world	45
3.4.1	Model 1: Minimal active inference	46
3.4.2	Model 2: Hierarchical active inference	51
3.4.3	Model 3: Counterfactual active inference	56
3.5	Two options for cognition	60
3.6	References	64
4	Embodiment in mind: The role of rhythmic visceral dynamics in cog- nitive development	75
4.1	Introduction	77
4.2	Getting a grip on embodied cognition	78
4.2.1	Varieties of embodiment	79
4.2.2	Causation versus constitution	80
4.3	Predictive processing and (embodied) active inference	81
4.3.1	Embodied models	83
4.3.2	Embodied feelings	84
4.3.3	Embodied selves	85
4.3.4	Embodied rhythms	86
4.4	A diachronic perspective on embodiment	88
4.4.1	Learning from within	90
4.4.2	Visceral afferent training drives activity-dependent neuronal de- velopment	93
4.4.3	Visceral afferent training inculcates a model of periodic fluctuation	94
4.4.4	Visceral afferent training signals promote self-organising brain dy- namics	96
4.5	Causation versus constitution redux	98
4.5.1	Constitution through causation?	98
4.5.2	Embodied mind or envatted brain?	100
4.6	Prospects for a unified philosophy of the embodied mind	102
4.7	Conclusion: The body as first teacher	103
5	Be still my heart: Cardiac regulation as a mode of uncertainty reduc- tion	104
5.0.1	Psychophysiology: The science of embodiment	105
5.0.2	Psychophysiology and the study of covert processes	107
5.0.3	Covert action under active inference	108
5.0.4	The present manuscript	110
5.1	Introduction	111
5.2	Materials and methods	113
5.2.1	Participants	113
5.2.2	Psychophysical stimuli and apparatus	113
5.2.3	Procedure	113

5.2.4	Electrophysiological signal acquisition and preprocessing	114
5.2.5	Data analysis	114
5.3	Results	115
5.3.1	Behavioral performance	115
5.3.2	Inter-beat interval and heart rate variability	115
5.3.3	Instantaneous heart frequency	115
5.3.4	Pulse wave and skin potential amplitude	116
5.4	Discussion	117
5.5	References	120
6	Restless hearts and wandering minds: The cardiac correlates of task-unrelated thought	124
6.0.1	Cardiac deceleration: From orientation to action	125
6.0.2	Selective attention, inhibition, and executive control	128
6.0.3	Performance monitoring and error processing	130
6.0.4	The present manuscript	132
6.1	Introduction	134
6.1.1	Mind-wandering methodology	134
6.1.2	The psychophysiology of mind-wandering	135
6.1.3	The psychophysiology of attention	136
6.1.4	The current study	137
6.2	Methods	138
6.2.1	Participants	138
6.2.2	Materials and apparatus	139
6.2.3	Procedure	139
6.2.4	Electrophysiological signal acquisition and preprocessing	141
6.2.5	Data analysis	141
6.2.5.1	Behavioural analysis	141
6.2.5.2	Thought probes	143
6.2.5.3	Cardiac parameters	143
6.2.5.4	Event-related inter-beat interval analysis	144
6.2.5.5	Cardiac cycle phase analysis	144
6.2.5.6	Model estimation, evaluation, and visualisation	145
6.2.6	Data availability statement	145
6.3	Results	145
6.3.1	Behavioural performance	145
6.3.1.1	Reaction time	146
6.3.1.2	Response accuracy	146
6.3.2	Attentional states	149
6.3.2.1	Behavioural performance during probe-defined epochs	150
6.3.3	Cardiac parameters	152
6.3.3.1	Cardiac parameters and behavioural performance	153
6.3.3.2	Cardiac parameters and attentional state	153
6.3.4	Event-related inter-beat interval analysis	154
6.3.5	Cardiac cycle phase analysis	155
6.4	Discussion	158
6.4.1	Time-on-task effects in the SART	159

6.4.2	Cardiac and cognitive control across time	160
6.4.3	Attention and behaviour in cardiac time	161
6.5	Conclusion	164
7	Finding meaning in the noise: Expectations guide attention towards the content of degraded speech	165
7.0.1	Language, hierarchy, and prediction	166
7.0.2	Predictive coding and predictive timing in speech perception . . .	168
7.0.3	The problem with predictive coding	170
7.0.4	The present manuscript	171
7.1	Introduction	174
7.1.1	Perceptual restoration of degraded speech	174
7.1.2	Neural mechanisms of perceptual filling-in	175
7.1.3	The current study	176
7.2	Methods	177
7.2.1	Participants	177
7.2.2	Stimuli	177
7.2.3	Procedure	178
7.2.4	EEG acquisition and preprocessing	179
7.2.5	Data analysis	180
7.2.5.1	Time-frequency decomposition	180
7.2.5.2	Stimulus reconstruction	181
7.2.5.3	Statistical analysis	182
7.3	Results	184
7.3.1	Correct prior information evokes perceptual pop-out	184
7.3.2	Prior knowledge exerts frequency-specific effects on sentence processing	184
7.3.3	Correct prior information enhances stimulus reconstruction	187
7.4	Discussion	188
7.4.1	Sentence pop-out is accompanied by enhanced stimulus reconstruction and theta suppression	188
7.4.2	Provision of prior information enhances delta- and alpha-band activity	191
7.5	Conclusion	192
8	Summary and concluding remarks	193
8.1	Looking back	193
8.2	Looking ahead	197
A	Supplementary materials	201
A.1	Materials: Face and house stimuli	203
A.2	Results	204
	Bibliography	212

1

Introduction

This is a thesis about life and mind. Its core ideas are rooted in a set of conceptual dichotomies that revolve (or rather, *oscillate*) around the central organising theme of *uncertainty*. These sometimes overlapping dichotomies include notions of stability and variability, difference and repetition, perception and action, and self and other, to name a few. Since each of the themes examined here are instantiated in – or contextualised by – some form of time-evolving process, this thesis is deeply concerned with the way certain kinds of events unfold through time. Although the concepts of time and process are not engaged with philosophically, the discussions that follow are in some important sense ‘shot through’ with temporality.

The key idea pursued in this thesis is that biological regulation serves to resolve uncertainty. The term ‘biological regulation’ is deliberately vague, ranging from relatively simple modulations of behaviour in single-celled organisms, to highly complex ensembles of co-ordinated neural activity in the primate brain. The term ‘resolve uncertainty’ is similarly vague, and will turn out to have various meanings depending on the specific context in play. One of the core tasks of this thesis is to delineate different kinds of biological regulation and uncertainty on the basis of formal principles – namely, those availed by the *free energy principle* and its corollaries.

1.1 The free energy principle

The free energy principle represents one of the most exciting and ambitious theoretical developments of the past 20 years. Early incarnations of the principle sought to provide a biologically-plausible, Bayesian-inspired account of the neural dynamics underwriting perceptual learning (Friston, 2002, 2005; Friston et al., 2006). Since then, its explanatory aspirations have grown considerably; in the hands of its chief architect, the free energy principle seems to hold the key to life, the universe, and every thing (Friston, 2013, 2019a). Unsurprisingly, the totalising ambitions of this project have invited a great deal of controversy, sparking lively philosophical debates about its metaphysical commitments, epistemic status, and scientific utility (see, e.g., Andrews 2021; Colombo and Wright 2018; Hohwy 2020b; Williams 2020).

While fascinating in their own right, such debates are not the concern of this thesis. Indeed, I mount no defence – nor attempt any criticism – of the free energy principle and its corollaries; rather, what is pursued here is an exploration of the insights that may be derived under its assumption. The overarching argument that unfolds over the course of this thesis thus takes a conditional form: If one commits to the free energy principle as a *first* principle, the implications for adaptive biological systems are *thus* and *so*.

The implications that are of greatest interest here are those pertaining to the nexus between physiological adaptation and cognitive function. Since these concepts reside at the very core of the free energy principle (as discussed in Chapter 2 and Chapter 3), I feel warranted in examining them while remaining largely agnostic about more recent attempts to extend its explanatory scope beyond the realm of self-organising biological systems. That said, I consider this conception of the free energy principle to be sufficiently general and substantive as to afford a genuinely interesting and original perspective on a range of issues in philosophy, theoretical biology, and cognitive neuroscience.

1.2 Interoceptive inference

A natural way to approach the topic of physiological regulation from the perspective of the free energy principle is via a family of theoretical accounts that fall under the rubric of *interoceptive inference* (Owens et al., 2018; Seth, 2013). Such accounts draw on process theories associated with the free energy principle (e.g., predictive coding, active

inference¹) to explain how the brain monitors and controls internal bodily states in order to maintain biological viability. Importantly, these theories also address the impact of physiological signals on brain dynamics and conscious awareness, thus emphasising the bidirectional, recurrent nature of brain-body communication.

Much of the work conducted in this arena has focused on the impact of interoceptive feedback on affective and self-related experience (Barrett and Simmons, 2015; Pezzulo, 2014; Quattrocki and Friston, 2014; Seth and Tsakiris, 2018), synthesising and developing on insights from the James-Lange (James, 1884; Lange, 1922), Cannon-Bard (Bard, 1928; Cannon, 1927), and Schachter-Singer (Schachter and Singer, 1962) theories of emotion. Another prominent line of research has sought to characterise different dimensions of interoceptive experience (Garfinkel et al., 2015), and how individual differences along these dimensions may modulate other domains of cognitive processing. Yet another recent trend in the literature is the resurgence of psychophysiological ‘cycle-timing’ studies, which examine how sensorimotor (or higher cognitive) processes vary as a function of their timing with respect to physiological cycles such as the heartbeat (Azzalini et al., 2019; Critchley and Garfinkel, 2018; Park and Tallon-Baudry, 2014).

These rich lines of inquiry are united by a common interest in the way internal bodily states influence the biological agent’s subjective experience of itself and its environment. The approach adopted here aims to complement the largely ‘bottom-up’ focus of such inquiry with a more explicitly ‘top-down’ account – one which construes the internal environment not only as a source of sensory information, but also as a site for proactive regulation or *allostasis*. This perspective encourages a broader, more temporally-extended (*diachronic*) view of the co-ordinated patterns of sensorimotor activity that unfold across interoceptive and exteroceptive domains. Furthermore, by eschewing experimental methods designed to induce affective states or direct attention towards body-related sensations, this approach may also unveil novel insights about the fundamentality of neuro-visceral dynamics in the cognitive economy at large (see Chapter 4).

1.3 Active sensing and covert action

A key motif within the active inference literature is the deeply reciprocal (i.e. ‘circular-causal’) interplay between perception and action. This idea is nicely captured by the concept of *active sensing* (or *active sampling*), which thematises the dependence of perceptual states on various sorts of action. This concept has been examined most

¹Strictly speaking, active inference isn’t a process theory in itself, but has a process theory associated with it (see Friston et al. 2017a). I will sometimes refer to ‘active inference’ rather loosely as a shorthand for this process theory.

extensively within the visual modality – specifically, the programming of saccadic eye movements that resolve uncertainty about the visual scene (Friston et al., 2012; Mirza et al., 2016; Parr and Friston, 2017). More recent work has extended this logic to the domain of *mental* action (Metzinger, 2017), whereby covert-attentional (rather than overt-motoric) actions are deployed to resolve ambiguity within sensory input (Friston et al., 2021; Parr et al., 2019).

Abstracting away from the specifics of any given sensory modality, the free energy principle implies that the agent should sample those actions it deems most likely to resolve uncertainty about evolving environmental dynamics. This perspective finesses classical cybernetic notions of feedback-driven control, in which actions are designed to maintain or reinstate expected sensory states. Active inference can thus be said to endow traditional control schemes with the capacity to seek out novel information (i.e. explore new regions of parameter space) that may confer some adaptive benefit in the future (see Chapter 3; see also Corcoran 2019). The notion of allostatic regulation sits comfortably with this idea, insofar as it licences deviations away from preferred reference or ‘setpoint’ states in order to optimise uncertainty over the long run.

As alluded to above, this thesis sets out to articulate the relation between interoceptive/autonomic state regulation and uncertainty reduction. Part of this story involves conceiving of interoceptive inference as a way of procuring sensory information in much the same way as exteroceptive channels are selectively sampled under regimes of active sensation. On this view, the functional contribution of interoceptive processing is not limited to the monitoring of internal bodily states, but extends to the optimisation of sensorimotor processing over multiple streams and scales. This idea brings into focus the crucial role of attention in arbitrating between alternate sources of sensory information, and by extension, in mediating a deep continuity between cognitive and physiological modes of regulation.

1.4 Overview of thesis structure

The main body of this thesis is divided into two halves, each of which comprise three chapters. The first set of chapters are predominantly concerned with theoretical matters, engaging in a philosophical analysis of biological regulation and its relation to cognition. With this theoretical groundwork in place, the second half of the thesis switches to an empirical mode. The main goal of these latter chapters is to present three psychophysiological studies that help to relate general principles of biological regulation with specific expressions of cognitive regulation. The contents of each chapter are briefly summarised below.

Chapter 2 introduces the concept of allostasis through the analysis of its historical development in the fields of physiology, biomedicine, and ethology. The nature of allostatic regulation, and its relation to more traditional modes of homeostatic control, are explored through three influential perspectives. These characterisations are then brought to bear on the various conceptions of allostasis that have arisen within the active inference literature. Although primarily concerned with understanding how these accounts cohere with or depart from pre-existing notions of allostasis, this analysis also raises some important philosophical questions about the scope and function of allostatic regulation within the broader scheme of active inference. These issues are revisited in greater detail over the subsequent two chapters.

Chapter 3 can be considered as a companion piece to Chapter 2 that reintroduces the concepts of homeostasis and allostasis from first principles. This chapter begins with a more detailed overview of the free energy principle and its corollaries, thereby laying the groundwork for a generative model-based account of the way allostatic mechanisms might be instantiated in complex biological organisms. In particular, the formal scheme availed by the free energy principle is used to analyse how different modes of representation and adaptation might arise from different sorts of computational architecture (i.e. generative model). This analysis motivates a principled distinction between allostatic and cognitive systems, whereby the latter are characterised in terms of their ability to engage in a counterfactual form of active inference.

Chapter 4 evaluates the philosophical significance of interoceptive inference for debates about embodied cognition. While many proponents of active inference believe it to have overcome longstanding disagreements between traditional cognitivism and more radical views about the embodied, embedded, and enactive nature of cognition, I argue that most of the empirical evidence adduced in favour of such perspectives offers little motivation for such claims. However, a novel perspective on the role of the body in shaping cognition is proposed based on a systems-level interpretation of active inference in the developing foetus.

Chapter 5 marks the turn towards empirical (specifically, *psychophysiological*) methods, and with it a thematic shift towards cognitive regulation and covert action. This chapter reports an binocular rivalry experiment investigating how heartbeat dynamics are adapted in response to different forms of visual uncertainty. This approach leverages recent theoretical developments in interoceptive/embodied active inference (Allen et al., 2019) to reconceptualise the psychophysiological correlates of attentive observation under a more general scheme of uncertainty reduction. Notably, this perspective reinvigorates early theoretical insights about the bidirectional nature of brain-heart communication (Lacey, 1959) that speak to the ideas raised in Chapter 4.

Chapter 6 builds on the findings of the previous chapter by moving beyond attentive observation to focus on other forms of cognitive control, including simple instances of action preparation, inhibition, and performance monitoring. A Sustained Attention to Response Task was combined with a thought-sampling methodology from the mind-wandering literature in order to gather data about the co-evolution of cardiac states, behavioural activity, and attentional dynamics over time. From a mind-wandering perspective, this study provides the first in-depth investigation of cardiac fluctuations during task-(un)related thought across multiple temporal scales. From an executive control perspective, this study extends previous psychophysiological investigations by assessing how behavioural-autonomic interactions vary under endogenous fluctuations of attentional state.

Chapter 7 departs from the previous three chapters' concern with bodily (and in particular, cardiac) rhythms, focusing instead on the challenges posed by another biologically salient form of rhythmic stimulation: continuous speech. This study investigated the effect of prior (written) information on one's ability to comprehend severely degraded spoken sentences. This paradigm was of particular interest not only because it presents another example of a perceptual recognition problem in the context of sensory ambiguity (cf. Chapter 5), but also in light of a recent active inference model characterising active listening as a form of covert mental action (Friston et al., 2021).

Chapter 8 concludes the thesis with a brief summary of the major outcomes of each chapter, and some considerations for future research.

Over the course of the following chapters, I hope to cast new light on the ways living creatures adapt to the vicissitudes of life. My starting point is the somewhat banal observation that biological existence is at once precarious and uncertain. It can be tempting to view the uncertainty of one's own life in the brilliant, variegated hues of its particularity: We find ourselves enmeshed within a complex matrix of physical, psychological, and social dynamics, each facing up to the specific challenges and existential threats of our time. Such details clearly matter. Nonetheless, there is much to be gained by examining complex phenomena under a more uniform glow – one that throws certain coarse-grained distinctions into relief. It is in this spirit that I proceed, leveraging the considerable power of the free energy principle to illuminate fundamental similarities and differences in the way cells, systems, and species deal with uncertainty.

2

Allostasis, interoception, and the free energy principle: Feeling our way forward

As alluded to in Chapter [1](#), the next two chapters lay out the conceptual groundwork on which much of this thesis rests. Both chapters are essentially concerned with the nature of allostasis, and its function within the broader scheme of biological regulation and adaptive action. The current chapter deals primarily with definitional matters, seeking to distill the core features of allostasis from its various characterisations in the literature. The chapter then addresses the way this concept has been deployed within more recent models of interoceptive processing and active inference. Although this latter discussion necessitates some exposition of the free energy principle and its attendant process theories, a more in-depth treatment of these topics will be postponed until Chapter [3](#).

Allostasis, interoception, and the free energy principle: Feeling our way forward

Andrew W. Corcoran and Jakob Hohwy

15.1 Introduction

The free energy principle (Friston, 2010) invokes variational Bayesian methods to explain how biological systems maximize evidence for their predictive models via the minimization of variational free energy, a tractable information-theoretic quantification of prediction error. This account, which was originally proposed to explain sensory learning, has evolved into a much broader scheme encompassing action and motor control, decision-making, attention, communication, and many other aspects of mental function (for overviews see Clark, 2013, 2016; Hohwy, 2013). Under the free energy principle, minimization of free energy is what any self-organizing system is compelled to do in order to resist dissipation and maximize the evidence for its own existence (i.e. self-evidencing through active inference; Hohwy, 2016).

Recent years have witnessed a growing interest in extending the conceptual apparatus of the free energy principle to the interoceptive domain. A number of investigators have sought to explain the influence of interoceptive modes of prediction error minimization on various cognitive processes and disruptions (for recent reviews see Barrett, 2017; Khalsa et al., in press; Seth & Friston, 2016, Smith et al., 2017). Central to such *interoceptive inference* perspectives is the notion that interoceptive signals encode representations of the internal (physiological) state of the body, thus providing vital information about how well the organism is managing to preserve the biological viability of its internal environment. Traditionally, the latter has been conceived in terms of *homeostasis*, a concept that usually refers (minimally) to the process of maintaining the internal conditions of complex, thermodynamically open, self-organizing biological systems in stable, far-from-equilibrium states (Yates, 1996). From the perspective of the free energy principle, homeostasis translates to the process of restricting the organism to visiting a relatively small number of states that are conducive to its ongoing existence, with interoceptive prediction error playing a particularly important role in signaling deviation from these attractive states (technically, these are known as attracting sets).

Notably, the centrality of homeostasis in some free energy-inspired accounts of interoceptive processing has started to give way to the newer concept of *allostasis*. According

to proponents of the latter, homeostasis fails to capture the rich variety of self-regulatory processes that biological systems engage in in order to conserve their own integrity. Allostasis tries to address this shortcoming through various theoretical innovations, chief amongst which is a core emphasis on predictive or anticipatory modes of regulation. This is to say that, rather than merely responding to physiological perturbations in order to ensure the internal conditions of the body remain within homeostatic bounds, allostasis enables the organism to proactively prepare for such disturbances *before* they occur.

While this account carries obvious appeal from the perspective of predictive model-based theories of interoceptive processing, attempts to marry the two have given rise to a number of divergent interpretations of allostatic regulation. As it turns out, the history of allostasis is a history of contested definitions; some 30 years on from its inception, there appears to be no definitive consensus as to its precise meaning. The key aims of this chapter, then, are to establish (a) how allostasis might be best understood as a distinctive concept in the overall scheme of biological regulation, and (b) how this construal might inform (and indeed, be informed by) free energy-inspired theories of interoceptive inference.

15.2 Discovering “the wisdom of the body”: Homeostasis

A standard account of the history of homeostasis might trace its source to the nineteenth century physiologist Claude Bernard, whose pioneering work on the role of the nervous system in maintaining the relative constancy of internal states (*le milieu intérieur*, i.e. the extracellular fluid environment that envelops the cell) would prove highly influential (Cooper, 2008; Woods & Ramsay, 2007). The key ideas at the core of Bernard’s thinking— notions of harmony, equilibrium, and regulation—are, however, much older, dating as far back as the pre-Socratics (see Adolph, 1961, for a historical review). Bernard refined the ancient insight that organisms maintain a healthy constitution by engaging in certain self-regulatory behaviors (e.g. consuming nutrients, excreting waste)—and deviate from well-being whenever subject to certain unfavorable physiological imbalances—by drawing attention to the physiological mechanisms that ensure the continuity of a stable internal environment. Such compensatory adjustments act to cancel out internal disturbances that would otherwise be caused by fluctuations in the external environment. This capacity to meet environmental impingements with countervailing responses thus grants the organism an adaptive coupling with—and a special kind of autonomy from—its environmental niche.

Benefitting from Bernard’s keen insights and some 50 years of subsequent experimental research, Walter Cannon (1929, 1939) coined the term “homeostasis” to describe the organism’s capacity to maintain a “steady state” or intrinsic uniformity despite ongoing fluctuations in its internal and external processes. Cannon was at pains, however, to stress that his neologism was intended to characterize a complex process in which multiple physiological mechanisms are recruited to ensure the continued stability of the organism’s internal milieu, where stability is construed in terms of a more or less variable range of acceptable (i.e. viable) values. This latter point is crucial for distinguishing

Cannon's conception of homeostasis from Bernard's emphasis on the fixed, invariant nature of internal conditions, in as much as homeostatic processes admit a space of permissible states. Also important was Cannon's concern to elucidate the autonomic mechanisms responsible for mediating adaptive physiological responses (e.g. increased respiratory rate) to altered internal conditions (e.g. decreased blood pH, increased carbon dioxide concentration) (Cooper, 2008), a subtle reorientation that would prove highly influential for later work in cybernetics.

The basic concept of homeostasis elaborated by Cannon (and extended by contemporaries such as Curt Richter; see Woods & Ramsay, 2007) would become one of, if not *the* core theoretical principle of modern physiology (Michael & McFarland, 2011; Michael et al., 2009). One essential element of modern conceptions of homeostasis that was, however, still missing from the Cannonian picture was a formal account of negative feedback (Modell et al., 2015). From a control-theoretic perspective, Cannon's careful analysis of particular homeostatic processes can be conceived according to a generic scheme of error detection (i.e. where some regulated variable, for instance blood glucose concentration, is found to deviate from some desirable value or *setpoint*) and correction (i.e. where some effector mechanism is activated in order to restore the regulated variable to the prescribed setpoint). It is important to note here that the notion of a setpoint generally conforms to Cannon's conception of a (broader or narrower) range of acceptable values, rather than any singular, fixed level (Modell et al., 2015). This set of values can thus be construed as a model against which the actual (sensed) state of the regulated variable is compared. The error signal elicited when the current state of the monitored tissue deviates from its setpoint reference represents a threat to organismic viability, and must therefore be corrected via mobilization of the appropriate effector system(s). Recasting homeostasis in this light thus furnishes a powerful conceptual framework in which the processes responsible for maintaining internal stability achieve this goal through the communication of information between peripheral tissues and a central controller (such as the central nervous system).

15.3 Allostasis: The future of homeostatic regulation?

As mentioned in our introduction, several recent theoretical frameworks of interoceptive inference co-opt notions of allostasis in order to situate the autonomic regulation of the internal milieu within the broader scheme of hierarchical predictive processing. Various theorists have argued that the basic concept of homeostasis is somehow insufficient to account for the rich complexity of self-regulatory behavior evinced by humans and other animals, advocating allostasis as a necessary theoretical supplement or corrective. To what extent allostasis extends, encompasses, or eliminates homeostasis is, however, unclear, not least because the characteristic features of allostatic regulation have been espoused in ambiguous or inconsistent terms across the literature (Lowe, Almér, & Dodig-Crnkovic, 2017; Power, 2004; Schulkin, 2004). This section thus aims to canvass some of the most influential accounts of allostasis to have emerged over the past three decades.

15.3.1 Achieving stability through change

The term “allostasis” was originally introduced by Sterling and Eyer (1988) to describe the integrated, hierarchical mechanisms through which the nervous system maintains organismic integrity. In this scheme, the brain is responsible for orchestrating complex, multisystem responses to physiological perturbations, resulting in a cascade of mutually reinforcing effects that are designed to maintain “stability through change” (Sterling & Eyer, 1988, p. 636). Multilevel allostatic regulation is supposedly accomplished through a fine-grained network of feedforward and feedback mechanisms, thus affording a more flexible and coordinated means of physiological control than the rather more primitive negative feedback loops typically attributed to homeostatic regulation. One key advantage of this arrangement is that it enables anticipatory alterations of physiological parameters *prior* to undergoing some perturbation (e.g. increasing blood pressure before standing up from a chair, rather than correcting the hypotension induced by the postural change after the fact). Under this allostatic regime, the body benefits from the brain’s capacity to learn from experience by forecasting the organism’s physiological needs ahead of time. As such, allostasis represents a rather more sophisticated system of internal regulation, one which minimizes reliance upon the kind of error signaling required to drive homeostatic correction.

Sterling and Eyer argued that the concept of homeostasis is fatally deficient, and ought thus to be “superseded” by their notion of allostasis (1988, p. 646; see also Sterling, 2004, 2012; Sterling & Laughlin, 2015). However, the validity of this assertion has been challenged by critics who argue that it turns on a fundamentally mistaken construal of homeostatic regulation (Carpenter, 2004; Day, 2005). The source of this error is twofold. First, the careful nuance of Cannon’s (1929, 1939) definition of homeostasis is ignored in this account, giving rise to the overly simplistic (and arguably misleading) impression that homeostasis is supposed to “clamp each internal parameter at a ‘setpoint’” (Sterling, 2004, p. 17), except in response to emergency (i.e. potentially life-threatening) situations. Second, Sterling and Eyer (1988) conflate the physiological variables that are the target of homeostatic regulation with the control mechanisms tasked with the job of maintaining such variables within acceptable bounds. The idea that physiological parameters such as blood pressure should fluctuate significantly throughout the day does not constitute a counterexample to the homeostatic model; rather, these fluctuations are in the service of homeostasis precisely insofar as they ensure that the vital constituents and properties of the fluid matrix (e.g. blood pH, oxygen tension) remain suitable for cell functioning. On this reading then, allostasis appears little more than “an unnecessary re-statement of the concept of homeostasis” (Day, 2005, p. 1196).

15.3.2 Allostatic means for homeostatic ends

Since Sterling and Eyer’s (1988) introduction of the concept, less radical versions of allostasis have been developed that seek to complement or extend the scope of homeostatic regulation, rather than reject it wholesale. Early work by McEwen, Schulkin, and

colleagues (McEwen & Stellar, 1993; Schulkin, McEwen, & Gold, 1994) embraced allostasis as a promising framework for studying complex relations between stress, behavior, and chronic disease, and set about developing the concept of *allostatic load* to account for the potentially deleterious consequences of resisting stressful stimuli. (Although an important dimension of the allostatic framework developed by McEwen and others, notions relating to allostatic load/overload will not be considered here—but see Peters, McEwen, & Friston, 2017).

As these theories matured, however, a more distinctive articulation of the base concept of allostasis started to emerge. McEwen began to conceive of allostasis as “an essential component of maintaining homeostasis” (1998, p. 37); where the latter is limited to “systems . . . that are truly essential for life” (2000b, p. 173). According to this view, allostasis describes “the process for actively maintaining homeostasis” (McEwen, 2000b, p. 173); or alternatively, “the means by which the body re-establishes homeostasis in the face of a challenge” (McEwen, 2000a, p. 25). In collaboration with Wingfield, McEwen’s notion of allostatic regulation was further expanded to include setpoint adjustments in anticipation of cyclical changes across various temporal scales (McEwen & Wingfield, 2003, 2010). This conceptual development highlighted the circadian modulation of homeostatic parameters implicit in Sterling and Eyer’s (1988) paradigmatic example of allostatic change (i.e. the diurnal variation of blood pressure upon which phasic modulations are superposed), while also extending the scope of allostatic processes to incorporate broader aspects of animal well-being, reproduction, and ontogenetic adaptation (e.g. seasonal variations in physiology and behavior in preparation for hibernation or migration).

McEwen concedes that his construal of allostasis might seem almost identical to broader conceptions of homeostasis, such as the view promulgated by Cannon (McEwen, 2000b, 2004; McEwen & Wingfield, 2003). He insists, however, that the notion of the “steady state” at the core of Cannonian homeostasis is inherently vague, insofar as it fails to delineate vital (homeostatic) systems from those mechanisms which work to maintain their stability. It is not entirely clear though why such a distinction ought to be desired, or indeed, if it is even coherent in the context of McEwen’s broader framework. Dallman (2003) argued that so-called allostatic systems do not manifest qualitatively distinct properties as compared to their homeostatic counterparts, on the basis that such systems are responsible for a great deal of essential physiological and behavioral functions. Indeed, it seems strange to claim that allostatic mechanisms are not equally essential to survival if such adaptive systems play a crucial role in enabling the organism to flee (or better yet, entirely avoid) a deadly predator, for example.

Although arguments of this sort might be blunted by a more charitable interpretation of the key idea underlying McEwen’s proposed distinction (namely, that allostatic systems accommodate large fluctuations precisely so that those physiological parameters which cannot tolerate such lability are not pushed beyond their narrow limits; e.g. McEwen, 1998), it seems plausible that significant enough deviations in allostatic systems should likewise prove fatal. Furthermore, cross-species analysis suggests that setpoint flexibility does not constitute a reliable indicator of the relative importance of a given physiological

parameter (see Boulos & Rosenwasser, 2004). Nevertheless, McEwen and Wingfield's (2003) thematization of the multiple layers of predictive regulation that unfold across the life cycle strikes us a valuable addition to the allostasis framework, one which we take to be a genuine departure from traditional notions of homeostasis.

15.3.3 Two modes of sustained viability

Another account of allostatic regulation that seeks to integrate (rather than replace) conventional notions of homeostatic control was put forth by Schulkin and colleagues (Power & Schulkin, 2012; Rosen & Schulkin, 2004; Schulkin, 2003a, 2003b). Schulkin (2003a, 2003b) credits Cannon's conception of homeostasis with greater scope and sophistication than Sterling and Eyer (1988), while maintaining that some kind of supplementary concept is necessary in order to capture the full gamut of regulatory strategies exhibited by complex organisms (Power & Schulkin, 2012; Schulkin, 2003b). Schulkin expounds a version of allostasis in which brain-driven regulatory mechanisms effect fluctuating physiological and psychological states in the absence of any clear setpoint boundary. In particular, anticipatory (feedforward) hormonal processes are posited to play a crucial role in the emergence of many appetitive, self-protective, and socially orientated motivational drives (Schulkin, 2003b, 2011), as well as explaining the affective valence of emotional experiences that accompany such states (Rosen & Schulkin, 2004). Schulkin and colleagues (Power & Schulkin, 2012; Rosen & Schulkin, 2004; Schulkin, 2003b, 2004) thus advocate a broad conception of biological regulation, one in which homeostasis and allostasis constitute equally important (yet functionally opponent) mechanisms for maintaining the biological viability of the internal milieu.

In some sense, we might regard Schulkin's framework as a kind of synthesis of prior allostatic concepts. It clearly inherits from Sterling and Eyer's (1988) original conception of allostasis, retaining as it does an explicit emphasis on the role of anticipatory physiological changes in efficient adaptation to environmental diversity. It also takes up McEwen and Wingfield's (2003) temporal expansion of the concept to account for longer-term adaptive changes in response to various ecological and life cycle contexts (Schulkin, 2003b, 2004). However, by balancing the homeostatic imperative to conserve stability with the allostatic impulse towards dynamic state transition, Schulkin and colleagues thematize the deeper continuity uniting these apparently contradictory concepts. At the heart of these regulatory principles is not so much the immediate influence they exert over target physiological parameters (i.e. internal constancy versus variability) but rather the overarching goal that these mechanisms dually subserve: namely, the ongoing survival and reproductive success (i.e. evolutionary fitness) of the organism (Power & Schulkin, 2012; Schulkin, 2004; see also Power, 2004).

This is not to say that the regulatory frameworks described by Sterling, McEwen, and others do not also ground the emergence of allostatic mechanisms in the selective advantages they confer. The point here, rather, is that sustained biological viability (rather than some other criterion such as internal stability) seems to us the most plausible target towards which physiological and behavioral regulatory mechanisms are striving. By these

lights, there is no inherent contradiction between homeostatic and allostatic principles; they are merely different routes to the same end.

15.4 Allostasis and interoceptive inference

The imperative to maintain biological viability over time is at the very core of the free energy principle (Friston, 2010). Briefly, this principle begins with the observation that living entities must “maintain their sensory states within physiological bounds,” and that they do so by engaging in actions which maintain the integrity of their structural and dynamical organization (Friston, 2013, pp. 1–2). This restates the cybernetic insight that biological organisms resist the tendency towards disorder wrought by variable external conditions (Ashby, 1947, 1962). The central element of the principle is that such self-preserving adaptation is achieved via environmental exchanges enabled by the minimization of free energy (or, under simplifying assumptions, the long-term average of prediction error; Friston, 2010). Under most accounts invoking the free energy principle, the process of maintaining the biological agent’s internal milieu within the limited subset of states conducive to its ongoing existence is that of homeostasis (where homeostasis is understood more precisely in terms of minimizing the free energy of internal state trajectories in order to avoid surprise, i.e. minimize prediction error; Friston, 2010).

The concept of allostasis started to infiltrate this picture in conjunction with remarks on the necessity of maintaining homeostasis for survival (e.g. Friston, 2012; Friston et al., 2014; Moran et al., 2014). Such comments typically invoked allostasis in the same breath as homeostasis, without offering any indication as to how the two terms might refer to differentiated aspects of biological regulation. To our knowledge, the first attempt at characterizing a substantive notion of allostasis as an independent mode of physiological regulation within the context of free energy minimization was made by Gu and FitzGerald (2014). In the short period that has elapsed since, a number of investigators have imported allostasis into their own free energy-inspired accounts of interoceptive inference. Much like the original development of allostasis in the biomedical and ethological literatures however, the precise nature of allostatic control in these schemes has been elaborated in various ways. The time is ripe then to take stock of this nascent body of research, both to establish its continuities with—and departures from—pre-existing notions of allostasis, and to assess which interpretation(s) of the concept seem most promising from the free energy perspective.

For convenience, we divide these recent allostatic treatments of interoceptive inference into three broad classes: *behavioral*, *teleological*, and *diachronic* (see Figure 15.1). This division is not meant to be taken as absolute; indeed, these accounts share many similarities by dint of their common theoretical origins.

15.4.1 Behavioral allostasis

In their commentary on Seth’s (2013) theory of interoceptive inference, Gu and FitzGerald argue that the scope of predictive interoceptive processing should be extended beyond

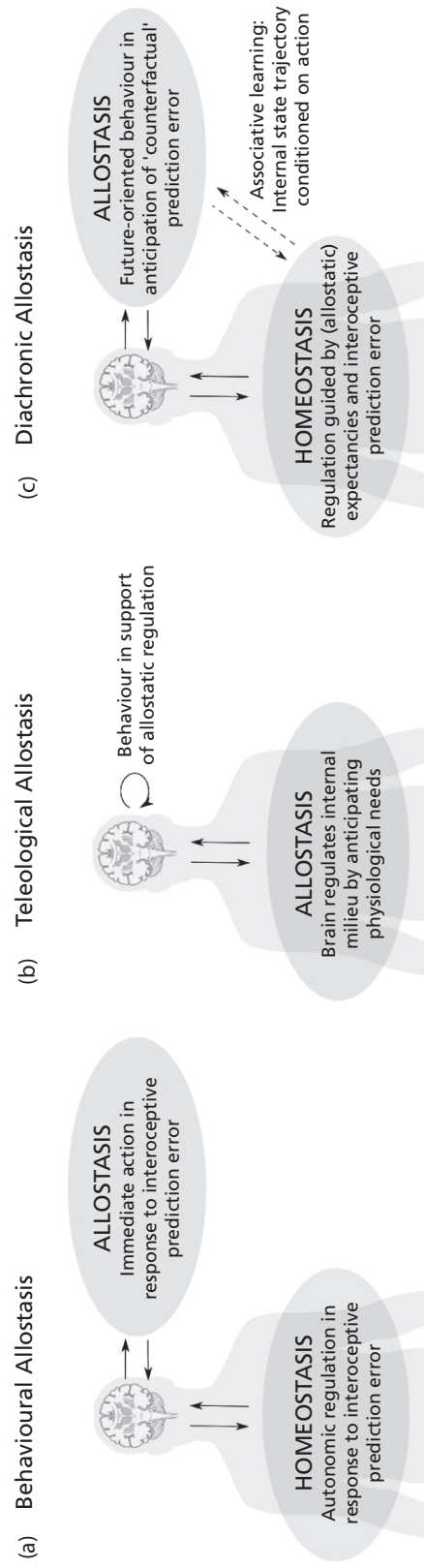


Figure 15.1 Schematic summary of the key conceptual distinctions between (a) behavioural, (b) teleological, and (c) diachronic accounts of allostatic regulation (see main text for details).

“homeostatic control of the internal milieu” to incorporate “allostatic actions on the external world” (2014, p. 269). At first, their position sounds isomorphic to that espoused by McEwen, insofar as allostasis is defined as “the process of achieving homeostasis” (Gu & FitzGerald, 2014, p. 269). It becomes quickly apparent, however, that Gu and FitzGerald (2014) conceive of homeostasis and allostasis in rather different terms. Here, homeostasis consists of autonomic reflexes that resist substantial fluctuations in the physiological conditions of the body (e.g. metabolizing stored fat in response to declining blood glucose levels), while allostasis corresponds to the behavioral actions that the agent undertakes in order to ameliorate some internal perturbation (e.g. consuming food in response to glucose decline). Gu and FitzGerald (2014) thus advocate a framework in which homeostatic (brain–internal world) and allostatic (brain–external world) loops offer alternative pathways to the same ultimate goal; namely, that of keeping the organism within the subset of biophysical states most conducive to its survival (in other words, minimizing the surprise or free energy indexed by interoceptive prediction error).

Gu and FitzGerald’s (2014) behavior-orientated characterization of allostasis is adopted and further elaborated by Seth (2015). Thematising the continuity between Ross Ashby’s pioneering work in cybernetics (Ashby, 1956, 1960) and the free energy principle, Seth (2015) seeks to map homeostasis and allostasis onto the “ultrastable” scheme exemplified by Ashby’s (1960) *homeostat*. Briefly, this device consists of four modular subsystems which dynamically interact to influence one another’s essential variables. If these interactions fail to preserve essential variables within an acceptable range, a regulatory switch intervenes to randomly reconfigure the system’s behavior. If the homeostat’s new organization fails to stabilize essential variables within range, it will continue to transition through its repertoire of possible configurations until stability is restored, or until the system disintegrates (see Cariani, 2009, for a more detailed explication of the homeostat’s functional architecture). Seth compares homeostasis to the first-order feedback loop constituted by the dynamic interplay of each module’s inputs and outputs, and allostasis to the second-order reorganization of these interactions (although allostatic behavior constitutes a purposeful, rather than random, attempt to transform system dynamics). On this account, then, allostatic behavior functions to alter the organism’s relation to its environment when homeostatic compensation fails to maintain physiological parameters within viable bounds.

Implementing these ideas within the context of free energy minimization, Seth (2015) argues that interoceptive prediction error can be minimized in one of three ways: (a) by adjusting model predictions in order to better approximate the incoming sensory signal (equivalently, updating one’s emotional state; i.e. perceptual inference); (b) by enlisting autonomic reflexes to alter internal conditions such that they correspond with the predicted internal state (i.e. active inference or first-order (homeostatic) control); or (c) by engaging in goal-directed behavior to act on the environment in such a way that brings about the predicted internal state (i.e. second-order (allostatic) control). Here, then, allostasis is not only distinguished from the physiological mechanisms responsible for regulating the internal milieu, but also construed as an alternative mode of achieving organismic viability.

An interesting aspect of Seth's (2015) analysis is the claim that perception simply "falls out" of the fundamental necessity to achieve homeostatic control. It is not entirely clear whether Seth subscribes to a kind of anti-realism which denies the veridicality of perceptual experience, or whether he wants to say that our rich perceptual experiences of the world are merely an accidental consequence of (or a useful tool for) the homeostatic imperative. In any case, Seth interprets the free energy principle in a way that assigns primacy to interoception (over exteroceptive perception), insofar as interoceptive inference is regarded as playing an instrumental role in steering the agent towards its homeostatic states. We shall encounter a similar view in section 15.4.2, hence we postpone further consideration of its implications until later. Let us first review the allostatic picture presented here.

Perhaps the most striking feature of these initial attempts to assimilate allostatic principles within a broader predictive processing framework is the surprisingly *reactive* way in which allostasis is depicted. Rather than presenting a paradigmatic example of *anticipatory* behavior in the service of some homeostatic goal (e.g. consuming food *prior* to the decline of blood glucose concentration), Gu and FitzGerald (2014) portray allostatic actions as a kind of external-world equivalent to the corrective autonomic responses orchestrated by homeostatic control mechanisms. Seth (2015) likewise articulates what seems to be a distinctly reactive form of allostatic regulation. Indeed, the ultrastable system to which Seth draws conceptual allusion is entirely dependent on negative feedback responses to the perturbation of essential variables. Although the second-order feedback loop is functionally analogous to McEwen's conception of allostasis as the means by which homeostatic variables are stabilized, this arrangement lacks the capacity to anticipate and offset such deviations before they occur (a vital feature of all allostatic frameworks reviewed in section 15.3). Consequently, the notion of allostasis invoked by these "behavioral" accounts does not obviously pick out any process that is distinctively predictive in nature.

Arguably, these fundamentally reactive models of allostasis derive from a partitioning of biological regulation along the lines of internal/autonomic (i.e. homeostatic) and external/goal-directed (i.e. allostatic) responses. Such a distinction is to our knowledge unprecedented in the allostasis literature, inasmuch as allostatic mechanisms have always been conceived as a suite of actions traversing the physiological—behavioral continuum. Here, notably, allostasis seems instead to refer exclusively to the behavioral strategies an agent can engage in response to mounting interoceptive prediction error, rather than a process that participates in the proactive avoidance of such surprising states. It is however unclear to us what substantive insights can be gleaned from this sort of picture. Indeed, it is so obvious that organisms must interact with their environment in order to satisfy their basic homeostatic needs (e.g. seeking out and drinking fluids to quench thirst) that such behavioral repertoires are a well-established feature of homeostatic theory (see e.g. Richter, 1942–43). Simply reassigning such activities under the rubric of allostasis is thus likely to revive the kind of criticism engendered by earlier renditions of the theory (e.g. that allostasis is essentially redundant insofar as it "represent[s] nothing that has not

always been part of the ordinary conceptual basis of homeostatic control,” Carpenter, 2004, p. 180).

On balance then, these interpretations risk diluting the concept of allostasis to the point where it constitutes little more than a particular mode of homeostasis, a behavioral rear-guard for occasions when autonomic mechanisms prove insufficient. As such, these inherently reactive accounts do not seem to carry us far beyond the insights availed by traditional homeostatic principles.

15.4.2 Teleological allostasis

The Embodied Predictive Interoception Coding model (EPIC; Barrett & Simmons, 2015) offers another free-energy inspired account grounding interoceptive experience in the physiological status of homeostatic variables. Initially, the authors of this model also defined allostasis in instrumental terms, describing it as the “process of activating physiological systems (such as hormonal, autonomic, or immune systems) with the aim of returning the body to homeostasis” (Barrett & Simmons, 2015, p. 422; Chanes & Barrett, 2016, p. 97). However, allostasis assumes a more pivotal role in subsequent work by Barrett and colleagues (Barrett, 2017; Barrett et al., 2016; Kleckner et al., 2017); the focus shifting from a reactive-mechanistic interpretation (i.e. where allostatic processes are recruited in response to homeostatic perturbation, similar to McEwen’s (1998, 2004) definition), to a broader perspective emphasizing its fundamentally predictive nature (i.e. where bodily conditions are efficiently regulated through the coordinated allocation of energy resources in anticipation of upcoming demands, similar to Sterling’s (2004, 2012; Sterling & Eyer, 1988) position). In this view, allostasis (and its interoceptive consequents) is assigned primary importance in the brain’s computational economy such that the predictive models posited to underpin cognitive representation are entirely subservient to the efficient satisfaction of the body’s physiological requirements (Barrett, 2017; Barrett et al., 2016).

Barrett and colleagues’ more recent characterizations of allostasis as the primary design feature driving brain evolution involves a number of important theoretical commitments. First, this expanded version of allostasis apparently subsumes the homeostatic functions that allostatic processes had previously been supposed to support. In eliminating all talk of homeostasis in favor of a more comprehensively encompassing model of predictive regulation, Barrett and colleagues (Barrett, 2017; Barrett et al., 2016; Kleckner et al., 2017) align themselves with Sterling’s (2004, 2012; Sterling & Laughlin, 2015) more radical allostatic agenda. It is not immediately clear that this sort of move is necessary for Barrett and colleagues’ more recent formulations to cohere, especially since their explicit concern with metabolic exchange and energy regulation would seem to sit just as comfortably within McEwen and Wingfield’s (2003) framework.

The second notable claim deriving from this framework is that the brain’s computational architecture has evolved in order to optimize allostatic regulation, rather than for purposes such as veridical perception or reasoned action (Barrett, 2017; Barrett et al.,

2016; Kleckner et al., 2017). This is to say that the brain's internal model (or "embodied simulation") of the body and the ecological niche it inhabits is fundamentally attuned to its physiological needs, such that only those features (i.e. statistical regularities) of the body–niche dyad relevant to allostatic regulation are represented (Barrett, 2017). Furthermore, Barrett and colleagues (Barrett, 2017; Barrett et al., 2016) propose that interoceptive representations emerge as a consequence of allostatic processing, and that such affective sensations form a fundamental and pervasive feature of conscious awareness. By implication, other sensory domains (and presumably, volitional motor activity) figure as secondary or derivative phenomena, the metabolic costs of which are tolerated only insofar as they furnish additional support to the brain's primary allostatic–interoceptive axis (Barrett, 2017).

This picture is reminiscent of Seth's (2015) argument for the primacy of interoceptive inference and physiological regulation. It is not entirely clear whether Barrett and colleagues consider higher-level cognitive functions to be useful adjuncts for maintaining allostasis, or whether they simply emerge as a byproduct of the brain's allostatic machinery. It is clear, however, that Barrett (2017) considers perceptual experience to be fundamentally driven by allostatic and interoceptive processing, such that one's subjective grasp of reality is modelled according to one's physiological needs. The upshot of this hypothesis is a constructivist account in which allostasis functions as the author and arbiter of phenomenological experience, both insofar as the imperative to optimize allostasis has carved out an evolutionary trajectory that has endowed the creature with a particular cognitive architecture and set of sensory capacities, and insofar as the experiential possibilities afforded by these devices are constrained and modulated in ways designed to realize this imperative in a given context.

The brain's evolution into a highly efficient allostatic machine, rather than (say) a rational decision-maker or accurate perceiver of the world, does not necessarily preclude the possibility that it should realize these additional properties also. Indeed, Seth, Barrett, and their colleagues may well agree that providing a creature with the capacity to accurately model the hidden causes of its external perturbations would, over the long-run, improve its capacity to maintain the viability of its internal milieu, as well as engage in other intrinsically rewarding (and evolutionarily relevant) projects such as reproductive activity. As Barrett (2017) points out, however, creatures need only be informed about hidden causes that are (potentially) relevant to their ongoing allostatic needs and priorities (for instance, evolution has endowed humans with a sensorium that is indifferent to infrared light stimulation). In this sense, then, these authors are correct to say that human perception does not afford a "true" picture of the world, at least insofar as the latter is construed as some complete account of the totality of measurable phenomena. Indeed, it is hard to imagine how the kind of experience that would obtain in the event that we really could perceive "everything" could be of much use, as dense with (predominantly irrelevant) information as it would be. There seems to be good *prima facie* reason then to think that (exteroceptive) sensation has evolved precisely to the extent that it is *useful*, and adaptive

self-regulatory activity (maximizing the likelihood of well-being and successful reproduction) would seem a reasonable object *for which* it ought to be useful.

These considerations notwithstanding, we note a general doubt about the plausibility of any thoroughgoing distinction between interoception and exteroception (independent of the specific role accorded to allostasis). Although it is true that the free energy principle allows for the possibility of inherited model parameters, and hence the newborn may come into the world equipped with certain expectations about the kinds of states its various sensory receptors ought to entertain, it is unclear why information conveyed via interoceptive afferents should be recognized by the brain as somehow different in kind to that received via exteroceptive (or proprioceptive) channels. From a brain-centric perspective, the external world to be modelled is that which lies beyond its neural projections, irrespective of whether this environment happens to be within or without the boundary formed by the body (Friston, 2010). In this respect, then, there is no meaningful distinction (for the brain) between the internal and external milieu; rather, there is only a Markov blanket (see Hohwy, 2017) separating a nervous system on the one side, and a hidden world of glucose molecules, blood vessels, muscles, fires, kittens, and so on, on the other. Collapsing this distinction leaves no principled rationale for privileging interoception over alternative forms sensory input; all channels furnish the brain with equally vital information about the state of play beyond the Markov blanket, from which its models profit.

A further, rather abstract concern about the teleological perspective presented here relates subtly to the conceptual role of the free energy principle. A key justification for the subordination of perceptual experience to homeostatic or allostatic regulation is made by way of appeal to the free energy principle's central concern with the persistent integrity of self-organized systems in the face of uncertain environmental conditions. Although we opened this section with a somewhat similar comment on the vital import of sustained biological viability in Friston's (2010) account, we urge caution in equating this with any so-called "fundamental imperative towards homeostasis" (Seth, 2015, p. 3). Rather, it would be more precise to say that the free energy principle captures something essential about the sorts of properties a biological system must possess in order to live (e.g. Friston & Stephan, 2007). It might be better then to say something like the following: any biological entity that consists of some form of sensorimotor interface through which it can enter into a dynamic exchange of energy and information with its environment, and which comprises an internal organization that enables it to minimize the free energy that bounds the surprise on its sensory states, is likely to endure; and in so doing, any such entity will thus *appear* to conform to the assumed imperative for the conservation of its biophysical integrity via self-regulatory processes. In other words, if a free energy-minimizing system exists, then it must indeed do so in virtue of possessing the right kind of internal configuration, and having entered into the right kind of circular-causal relationship with its environment, to be able to model the causes of its sensory states and engage in (what will look like) adaptive, self-regulatory activity (cf. Allen & Friston, 2018). As such, the apparent imperative towards self-regulatory behavior (be it homeostatic, allostatic, or whatever)

seems to fall out of the ongoing minimization of free energy in much the same way as the apparent teleological force driving evolutionary “design” emerges as a consequence of the intricate, non-teleological dynamics driving natural selection.

15.4.3 Diachronic allostasis

We turn finally to two remaining inferential formulations of allostasis, which we refer to as “diachronic” on account of the important implications they have for regulatory activity over various timescales.

Pezzulo, Rigoli, and Friston (2015) set out to explain how prospective and goal-directed (i.e. allostatic) forms of control might have evolved from more primitive mechanisms subserving homeostatic regulation. Here, homeostasis is construed along control-theoretic/cybernetic lines of negative feedback and setpoint control, where autonomic and behavioral reflexes are enlisted to correct deviations in physiological variables (see also Pezzulo, 2013; Seth, 2013). By contrast, allostasis refers to the flexible, context-specific engagement of complex, adaptive behavioral repertoires for the purposes of achieving some future outcome. Like the accounts surveyed in section 15.4.1, then, homeostasis and allostasis are equated with “direct” and “indirect” modes of eliminating interoceptive prediction error, respectively. Note however that the distinction here is more nuanced, insofar as homeostatic responses extend to the innate behavioral sets (e.g. approach/avoidance behavior) that equip animals to survive in the absence of associative learning.

If complex behavioral policies are to offer an effective means of controlling the physiological conditions of the body, it is essential that they deliver the right kinds of state transitions at the right time. This requirement is inherently challenging, however, since the consequences of a particular policy are necessarily realized some time after those conditions that triggered its initiation. Such delays are nontrivial in the context of homeostatic control, where a process causing physiological conditions to deteriorate may precipitate catastrophic damage if not promptly addressed. Pezzulo, Rigoli, and Friston’s (2015) solution to this problem leverages the free energy minimizing agent’s ability to acquire sophisticated internal models of the hidden environmental causes of its sensory states. Specifically, they argue that such generative interoceptive models enable such agents to predict the temporal evolution of interoceptive state trajectories (i.e. how interoceptive signals are likely to change over time), and encode how these trajectories correlate with sensorimotor events in the external world (cf. Friston et al., 2017). In virtue of the higher-level integration of sensory information converging from interoceptive, exteroceptive, and proprioceptive streams, the agent is thus able to acquire a rich understanding of how behavioral activities come to influence interoceptive states across various contexts. By linking interoceptive prediction errors and their suppression through active inference (i.e. engagement of allostatic behavior) via such associative learning processes, Pezzulo and colleagues (2015) provide a compelling explanation of (a) how the allostatic anticipation of future homeostatic needs might systematically arise, and (b) why allostatic behavioral policies should be endorsed despite potentially lengthy delays in their homeostatic payoff.

On this construal, allostatic processing turns out to be fundamentally *counterfactual* in nature. Higher (or deeper) hierarchical representations map the relation between increasingly distal outcome states and the behavioral policies that would lead towards their accomplishment. This account thus renders a smooth continuum of adaptive action selection, ranging from the primitive drives that work, for instance, to sate appetite via exploitation of the immediate environment, to the complex deliberative activities serving various motivations extending well beyond the basic requirements of the internal milieu (see also Pezzulo, 2017). Indeed, Pezzulo and colleagues (2015) observe that the capacity to learn the counterfactual relations that enable the agent to engage in prospective planning, and to choose amongst various available policies, confers an unparalleled degree of autonomy from the exigencies of the homeostatic imperative. Thus, in much the same way as Bernard and Cannon recognized how the capacity to maintain the stability of the internal milieu granted complex biological systems a remarkable degree of autonomy from the caprices of their external environments, allostasis under this scheme extends such freedom even further. Capable of holding the immediate demands of homeostasis in abeyance to some supraordinate desired (i.e. unsurprising and attracting) state, the autonomous horizon of the allostatic organism expands beyond the conditions of the present into a predictable (albeit uncertain) future.

Finally, Stephan, Manjaly, Mathis, and colleagues (2016) propose a formalized Bayesian implementation of hierarchical allostatic control that likewise operates across various temporal grains. Allostasis is defined here as the mode of active inference which performs “anticipatory homeostatic control” (Stephan et al., 2016, p. 5). This is achieved via the modulation of prior beliefs concerning the expected state trajectory of a given homeostatic setpoint. Expectancies about setpoint values are construed in terms of a probability distribution, such that beliefs propagated from higher-level circuits influence both the mean value of the controlled variable, and its associated variability (or precision). In other words, the traditional notion of a homeostatic negative feedback loop is situated at the lowest level of the processing hierarchy, with its target setpoint (i.e. the expected physiological state) conditioned by top-down information received from higher (allostatic) circuits. These higher (or deeper) hierarchical levels are posited to model increasingly broader, domain-general representations of the present state of the body and its environment, as well as predictions about changes in those states. Consequently, this account of allostatic regulation incorporates an important temporal dimension, where higher-level generative models are able to inform and update lower-level homeostatic control mechanisms in accordance with predictions about upcoming state transitions.

Stephan and colleagues (2016) set out their model of allostatically regulated homeostatic reflexes in accordance with the basic computational architecture assumed by the free energy principle. Homeostatic control thus depends on both the perception of salient features within the internal and external milieu (comprising both physical and social dynamics), and selection of appropriate actions designed to prevent dangerous (i.e. surprising) deviations of physiological parameters. Inference is divided into interoceptive and exteroceptive sensory processing. Prediction concerns how internal and external

states will evolve over time, as well as the degree to which possible actions will maintain internal states within the bounds of a given homeostatic setpoint over time. In other words, allostatic prior beliefs set expectations about the space of bodily states that the organism ought to inhabit (i.e. that delimited set of attracting states which engender low entropy), which homeostatic systems subsequently attempt to realize. Importantly, this generic active inference scheme is extended beyond the context of low-level homeostatic reflexes to encompass the higher-level implementation of flexible behavioral policies designed to avoid homeostatic surprise (in a similar vein to Pezzulo et al., 2015).

Stephan and colleagues (2016) present the first mathematically concrete account of allostatic control within the context of free energy minimization. Although more work needs to be done to flesh out this formal scheme with respect to the complex dynamics involved in the integrated regulation of complex physiological systems, it provides a plausible theoretical framework for explaining a number of core allostatic phenomena. The notion of a Bayesian reflex arc whose setpoint is adaptively defined and constrained by higher-order (allostatic) dynamics provides an elegant explanation of setpoint variability; one that seems equally capable of incorporating other (i.e. non-allostatic) accounts of flexible setpoint control (e.g. Cabanac, 2006). Embedding this arc within a hierarchical architecture also provides a principled mechanistic explanation of how certain higher-order parameters might be prioritized at the expense of less-urgent homeostatic needs, and how maladaptive psychological states might be entrained by persistent interoceptive prediction error. This perspective thus offers a deeply unifying picture of homeostatic and allostatic control as a dynamic coupling or closed loop, with lower-level homeostatic inferences and higher-level allostatic predictions reciprocally informing and modulating one another as the joint conditions of the agent–niche dyad evolve.

Aside from some minor technicalities concerning the precise definitional boundaries of homeostatic and allostatic control, we consider the diachronic theories reviewed in this section to be broadly compatible and complementary. We prefer Stephan and colleagues' (2016) Bayesian reflex formulation insofar as it expands the scope of allostatic control to the modulation of internal conditions (rather than limiting it to the domain of external, goal-directed behavior). This perspective is more consistent with the historical development of the allostatic framework (as examined in section 15.3), all prominent versions of which assume allostasis to consist of a repertoire of mechanisms that include the capacity to influence internal conditions directly by harnessing physiological effectors. Happily, the Bayesian reflex account invokes a principled distinction between homeostatic and allostatic control which succeeds in preserving the key functional characteristics of both modes of regulation (i.e. it neither collapses one concept into the other, nor relies on arbitrary or vague criteria for distinguishing their respective remits), while still allowing for the kind of higher-level, temporally-extended allostatic behavior articulated by Pezzulo and colleagues (2015). Furthermore, we find Stephan and colleagues' (2016) framework a potentially more useful starting point for future inquiry into the general nature of biological regulation, insofar as it affords the basic computational elements for scaffolding the emergence of less flexible, non-counterfactual forms of allostatic regulation (e.g.

circadian, circannual, and ontogenetic). By integrating the complementary perspectives provided by both diachronic theories, we arrive at a nuanced and fecund account of self-regulation that accommodates multiple scales of biological and cognitive complexity.

15.5 The future of the history of allostasis

Our review of the origins of allostasis, and analysis of its recent uptake in theories of interoceptive inference, might give the impression that the concept is as protean as the phenomena which inspired its coinage. This may be a consequence of zealous category splitting on our part, motivated by our intent to differentiate meaningful distinctions amongst a cluster of intersecting (and not entirely consistent) theoretical perspectives. However, the various interpretations and treatments allostasis has received over the years have tended to congeal around a more or less stable core of organizing principles (e.g. Schulkin, 2004). Mature versions of Sterling's (2004, 2012) and McEwen's (e.g. 2004, 2007) frameworks have understandably evolved into more expansive and nuanced iterations of their progenitors, benefitting from empirical advances and critical discussion. These influential accounts have thus reached a point of quasi-consensus, in as much as they lack the diversity of a genuine pluralism, but fail to converge fully on a coherent, unified account of what allostasis is or does. This leaves us in the somewhat precarious position of possessing a theoretical construct that appears well established and valid, but comprises a heterogeneous and not entirely coherent set of commitments. Part of the motivation of this chapter was therefore to highlight this situation, given that free energy theorists have started helping themselves to aspects of the allostasis construct without necessarily being explicit about which particular interpretation(s) of it they wish to endorse.

A useful illustration might be drawn from our distinction between what we dubbed the teleological and the diachronic interpretations of allostasis. Indeed, those familiar with the former might protest that it too invokes a hierarchical architecture which, much like the diachronic accounts, also admits of higher generative models encoding predictions extending across increasingly extended temporal windows. As such, it might seem somewhat disingenuous to exclude this model from our favored diachronic category. Our point, however, is that these frameworks are founded on rather different understandings of allostasis, giving rise to subtle but deep conceptual disagreements. The teleological perspective considers exteroception as secondary to interoception, which in turn emerges as a consequence of allostasis. The diachronic perspectives, on the other hand, seem to hold each domain of sensory information in equal standing; interoception, exteroception, and proprioception are blended together at a suitably high level of hierarchical modelling and without any indication that any stream is more fundamental than the others.

We urge care about which aspects of allostatic theory are imported into predictive model-based accounts of interoception. Indeed, it is notable that none of the interoceptive inference theories reviewed in this chapter acknowledge the accusations of redundancy, inconsistency, and ambiguity that have been levelled against the allostasis literature, even after some of these authors had substantially revised their own application of the concept.

Ignoring such issues not only belies the contested nature of allostatic control, it has the potential to propagate further confusion as disparate elements of the construct are selectively sampled and fused together.

If the future of allostasis is to disclose meaningful theoretical insights concerning the predictive processes that support biological regulation and interoceptive inference, then the next phase of its conceptual development requires us to work out a clear and precise understanding of its core principles and entailments. We have tried to clarify some of the confusion that has plagued the allostasis literature since its inception, and argued in favor of an inclusive view that reconciles homeostasis and allostasis as complementary strategies for sustaining biological viability. We have also attempted to shed light on some of the idiosyncratic ways in which allostasis has been deployed in recent characterizations of interoceptive inference, and suggest that future progress in this line of research will be hindered if these conceptual inconsistencies are not subject to critical scrutiny.

Acknowledgments

We thank the anonymous reviewer for their suggested improvements to an earlier version of this chapter. AWC is supported by an Australian Government Research Training Program (RTP) scholarship. JH is supported by The Australian Research Council DP160102770 and by the Research School Bochum and the Center for Mind, Brain and Cognitive Evolution, Ruhr-University Bochum. Author ORCIDs are as follows: AWC is 0000-0002-0449-4883; JH is 0000-0003-3906-3060.

References

- Adolph, E. F. (1961). Early concepts of physiological regulations. *Physiological Reviews*, 41(4), 737–70.
- Allen, M. and Friston, K. J. (2018). From cognitivism to autopoiesis: Towards a computational framework for the embodied mind. *Synthese*, 195(6), 2459–82.
- Ashby, W. R. (1947). The nervous system as physical machine: With special reference to the origin of adaptive behavior. *Mind*, 56(221), 44–59.
- Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Chapman & Hall Ltd.
- Ashby, W. R. (1960). *Design for a Brain: The Origin of Adaptive Behaviour*, 2nd edn. London: Chapman & Hall Ltd.
- Ashby, W. R. (1962). Principles of the self-organizing system. In: H. Von Foerster and G. W. Zopf Jr. (eds), *Principles of Self-Organization: Transactions of the University of Illinois Symposium*. London: Pergamon Press, pp. 255–78.
- Barrett, L. F. and Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, 16(7), 419–29.
- Barrett, L. F., Quigley, K. S., and Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B*, 371(20160011), 1–17.
- Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive & Affective Neuroscience*, 12(1), 1–23.
- Boulos, Z. and Rosenwasser, A. M. (2004). A chronobiological perspective on allostasis and its application to shift work. In: J. Schulkin (ed.), *Allostasis, Homeostasis, and the Costs of Physiological Adaptation*. Cambridge: Cambridge University Press, pp. 228–301.

- Cabanac, M. (2006). Adjustable set point: To honor Harold T. Hammel. *Journal of Applied Physiology*, 100(4), 1338–46.
- Cannon, W. B. (1929). Organization for physiological homeostasis. *Physiological Reviews*, 9(3), 399–431.
- Cannon, W. B. (1939). *The Wisdom of the Body*. New York, NY: W. W. Norton & Company, Inc.
- Cariani, P. A. (2009). The homeostat as embodiment of adaptive control. *International Journal of General Systems*, 38(2), 139–54.
- Carpenter, R. H. S. (2004). Homeostasis: A plea for a unified approach. *Advances in Physiology Education*, 28, 180–7.
- Chanes, L. and Barrett, L. F. (2016). Redefining the role of limbic areas in cortical processing. *Trends in Cognitive Sciences*, 20(2), 96–106.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral Brain Sciences*, 36(3), 181–253.
- Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.
- Cooper, S. J. (2008). From Claude Bernard to Walter Cannon. Emergence of the concept of homeostasis. *Appetite*, 51(3), 419–27.
- Dallman, M. F. (2003). Stress by any other name? *Hormones & Behavior*, 43(1), 18–20.
- Day, T. A. (2005). Defining stress as a prelude to mapping its neurocircuitry: No help from allostasis. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 29(8), 1195–200.
- Friston, K. J. and Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, 159(3), 417–58.
- Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–38.
- Friston, K. J. (2012). Embodied inference and spatial cognition. *Cognitive Processing*, 13(Suppl. 1) S171–77.
- Friston, K. J. (2013). Life as we know it. *Journal of the Royal Society Interface*, 10(86), 20130475.
- Friston, K. J., Rosch, R., Parr, T., Price, C., and Bowman, H. (2017). Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews*, 77, 388–402.
- Friston, K. J., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., and Dolan, R. J. (2014). The anatomy of choice: Dopamine and decision-making. *Philosophical Transactions of the Royal Society B*, 369(20130481), 1–12.
- Gu, X. and FitzGerald, T. H. B. (2014). Interoceptive inference: Homeostasis and decision-making. *Trends in Cognitive Sciences*, 18(6), 269–70.
- Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press.
- Hohwy, J. (2016). The self-evidencing brain. *Noûs*, 50(2), 259–85.
- Hohwy, J. (2017). How to entrain your evil demon. In: T. Metzinger and W. Wiese (eds), *Philosophy and Predictive Processing*. Frankfurt am Main: MIND Group, pp. 1–15.
- Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, J. S., Feinstein, J. S., et al. (in press). Interoception and mental health: A roadmap. *Biological Psychiatry: Cognitive Neuroscience & Neuroimaging*.
- Kleckner, I. R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W. K., et al. (2017). Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nature Human Behaviour*, 1(0069), 1–14.
- Lowe, R., Almér, A., and Dodig-Crnkovic, G. (2017). Predictive regulation in affective and adaptive behaviour: An allostatic-cybernetics perspective. In: J. Vallverdú, M. Mazzara, M. Talanov, S. Distefano, and R. Lowe (eds), *Advanced Research on Biologically Inspired Cognitive Architectures*. Hershey, PA: IGI Global, pp. 148–77.

- McEwen, B. S. and Stellar, E. (1993). Stress and the individual: Mechanisms leading to disease. *Archives of Internal Medicine*, 153(18), 2093–101.
- McEwen, B. S. (1998). Stress, adaptation, and disease: Allostasis and allostatic load. *Annals of the New York Academy of Sciences*, 840, 33–44.
- McEwen, B. S. (2000a). Protective and damaging effects of stress mediators: Central role of the brain. In: E. A. Mayer and C. B. Saper (eds), *Progress in Brain Research*, Vol. 122. Amsterdam: Elsevier Science, pp. 25–34.
- McEwen, B. S. (2000b). The neurobiology of stress: From serendipity to clinical relevance. *Brain Research*, 886(1–2), 172–89.
- McEwen, B. S. and Wingfield, J. C. (2003). The concept of allostasis in biology and biomedicine. *Hormones & Behavior*, 43(1), 2–15.
- McEwen, B. S. (2004). Protective and damaging effects of mediators of stress: Allostasis and allostatic load. In: J. Schulkin (ed.), *Allostasis, Homeostasis, and the Costs of Physiological Adaptation*. Cambridge, MA: MIT Press, pp. 65–98.
- McEwen, B. S. (2007). Physiology and neurobiology of stress and adaptation: Central role of the brain. *Physiological Reviews*, 87(3), 873–904.
- McEwen, B. S. and Wingfield, J. C. (2010). What is in a name? Integrating homeostasis, allostasis and stress. *Hormones & Behavior*, 57(2), 105–11.
- Michael, J., Modell, H., McFarland, J., and Cliff, W. (2009). The “core principles” of physiology: What should students understand? *Advances in Physiology Education*, 33(1), 10–16.
- Michael, J. and McFarland, J. (2011). The core principles (“big ideas”) of physiology: Results of faculty surveys. *Advances in Physiology Education*, 35(4), 336–341.
- Modell, H., Cliff, W., Michael, J., McFarland, J., Wenderoth, M. P., and Wright, A. (2015). A physiologist's view of homeostasis. *Advances in Physiology Education*, 39(4), 259–66.
- Moran, R. J., Symmonds, M., Dolan, R. J., and Friston, K. J. (2014). The brain ages optimally to model its environment: Evidence from sensory learning over the adult lifespan. *PLoS Computational Biology*, 10(1), e1003422.
- Peters, A., McEwen, B. S., and Friston, K. J. (2017). Uncertainty and stress: Why it causes diseases and how it is mastered by the brain. *Progress in Neurobiology*, 156, 164–88.
- Pezzulo, G. (2013). Why do you fear the bogeyman? An embodied predictive coding model of perceptual inference. *Cognitive, Affective, & Behavioral Neuroscience*, 14(3), 902–11.
- Pezzulo, G., Rigoli, F., and Friston, K. J. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, 134, 17–35.
- Pezzulo, G. (2017). Tracing the roots of cognition in predictive processing. In: T. Metzinger and W. Wiese (eds), *Philosophy and Predictive Processing*. Frankfurt am Main: MIND Group, pp. 1–20.
- Power, M. L. (2004). Commentary: Viability as opposed to stability: An evolutionary perspective on physiological regulation. In: J. Schulkin (ed.), *Allostasis, Homeostasis, and the Costs of Physiological Adaptation*. Cambridge: Cambridge University Press, pp. 343–64.
- Power, M. L. and Schulkin, J. (2012). Maternal obesity, metabolic disease, and allostatic load. *Physiology & Behavior*, 106(1), 22–8.
- Richter, C. P. (1942–43). Total self-regulatory functions in animals and human beings. *Harvey Lecture Series*, 38, 63–103.
- Rosen, J. B. and Schulkin, J. (2004). Adaptive fear, allostasis, and the pathology of anxiety and depression. In: J. Schulkin (ed.), *Allostasis, Homeostasis, and the Costs of Physiological Adaptation*. Cambridge: Cambridge University Press, pp. 164–227.
- Schulkin, J., McEwen, B. S., and Gold, P. W. (1994). Allostasis, amygdala, and anticipatory angst. *Neuroscience & Biobehavioral Reviews*, 18(3), 385–96.

- Schulkin, J. (2003a). Allostasis: A neural behavioral perspective. *Hormones & Behavior*, 43(1), 21–7.
- Schulkin, J. (2003b). *Rethinking Homeostasis: Allostatic Regulation in Physiology and Pathophysiology*. Cambridge, MA: MIT Press.
- Schulkin, J. (2004). Introduction. In: J. Schulkin (ed.), *Allostasis, Homeostasis, and the Costs of Physiological Adaptation*. Cambridge: Cambridge University Press, pp. 1–16.
- Schulkin, J. (2011). Social allostasis: Anticipatory regulation of the internal milieu. *Frontiers in Evolutionary Neuroscience*, 2(111), 1–15.
- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11), 565–73.
- Seth, A. K. (2015). The cybernetic Bayesian brain: From interoceptive inference to sensorimotor contingencies. In: T. Metzinger and J. M. Windt (eds), *Open Mind*. Frankfurt am Main: MIND Group, pp. 1–24.
- Seth, A. K. and Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B*, 371(1708), 1–10.
- Smith, R., Thayer, J. F., Khalsa, S. S., and Lane, R. D. (2017). The hierarchical basis of neurovisceral integration. *Neuroscience & Biobehavioral Reviews*, 75, 274–96.
- Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A. E., Paliwal, S., Gard, T., et al. (2016). Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience*, 10(550), 1–27.
- Sterling, P. and Eyer, J. (1988). Allostasis: A new paradigm to explain arousal pathology. In: S. Fisher and J. Reason (eds), *Handbook of Life Stress, Cognition and Health*. John Wiley & Sons Ltd, pp. 629–49.
- Sterling, P. (2004). Principles of allostasis: Optimal design, predictive regulation, pathophysiology and rational therapeutics. In: J. Schulkin (ed.), *Allostasis, Homeostasis, and the Costs of Physiological Adaptation*. Cambridge: Cambridge University Press, pp. 17–64.
- Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior*, 106(1), 5–15.
- Sterling, P. and Laughlin, S. (2015). *Principles of Neural Design*. Cambridge, MA: MIT Press.
- Woods, S. C. and Ramsay, D. S. (2007). Homeostasis: Beyond Curt Richter. *Appetite*, 49(2), 388–98.
- Yates, F. E. (1996). Homeostasis. In: J. E. Birren (ed.), *Encyclopedia of Gerontology: Age, Aging, and the Aged*, Vol. 1. San Diego, CA: Academic Press, pp. 679–86).

3

From allostatic agents to counterfactual cognisers: Active inference, biological regulation, and the origins of cognition

The previous chapter presented a broadly historical analysis that traced the emergence of allostasis from its roots in homeostatic and cybernetic control theory to its more recent assimilation within the active inference framework. The present chapter presents a complementary analysis that traces the emergence of allostatic dynamics from ‘first principles’; namely, the formal machinery of the free energy principle, its corollaries and process theories.

A major goal of this chapter is to show how the conceptual framework availed by active inference can inform and finesse philosophical debates about the function of cognition. Noting certain formal similarities between the role of uncertainty within the free energy principle and Peter Godfrey-Smith’s (1996) influential environmental complexity thesis, differences in the computational architecture underwriting adaptive behaviour are argued to distinguish allostatic from cognitive modes of regulation. This analysis thus helps to clarify the elision of physiological and more explicitly cognitive forms of prospective action noted in Chapter 2, thereby casting new light on the relation between life and mind.



From allostatic agents to counterfactual cognisers: active inference, biological regulation, and the origins of cognition

Andrew W. Corcoran¹ · Giovanni Pezzulo² · Jakob Hohwy¹

Received: 6 November 2019 / Accepted: 1 April 2020
© Springer Nature B.V. 2020

Abstract

What is the function of cognition? On one influential account, cognition evolved to co-ordinate behaviour with environmental change or *complexity* (Godfrey-Smith in *Complexity and the function of mind in nature*, Cambridge Studies in Philosophy and Biology, Cambridge University Press, Cambridge, 1996). Liberal interpretations of this view ascribe cognition to an extraordinarily broad set of biological systems—even bacteria, which modulate their activity in response to salient external cues, would seem to qualify as cognitive agents. However, equating cognition with adaptive flexibility per se glosses over important distinctions in the way biological organisms deal with environmental complexity. Drawing on contemporary advances in theoretical biology and computational neuroscience, we cash these distinctions out in terms of different kinds of generative models, and the representational and uncertainty-resolving capacities they afford. This analysis leads us to propose a formal criterion for delineating cognition from other, more pervasive forms of adaptive plasticity. On this view, biological cognition is rooted in a particular kind of functional organisation; namely, that which enables the agent to detach from the present and engage in counterfactual (active) inference.

Keywords Complexity · Uncertainty · Cognition · Allostasis · Homeostasis · Free energy principle · Active inference · Environmental complexity thesis · Adaptation · Representation · Interoception · Biorhythms · Life-mind continuity

✉ Andrew W. Corcoran
andrew.corcoran1@monash.edu

¹ Cognition and Philosophy Laboratory, School of Philosophical, Historical, and International Studies, Monash University, Melbourne, Australia

² Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche, Rome, Italy

Introduction

What is cognition? What is it *for*? While the former question is a perennial source of philosophical dispute, the latter seems to attract rather less controversy. Cognition—whatever it consists in and however realised—is ultimately functional to adaptive success. It enables the organism to register information about the state of its environment, and to exploit such information in the service of adaptive behaviour. Cognition, in short, is for *action*.

As benign as this characterisation might appear on first blush, a host of thornier questions lie in wait: Are *all* varieties of adaptive behaviour mediated by cognition, or only a select few? If the former, does this notion of behaviour extend to artificial and multi-agent systems, or is it limited to individual organisms? If the latter, what properties distinguish cognitive from non-cognitive modes of behaviour (assuming there *is* a clear distinction to be made)? And what of those cognitive processes that seem entirely encapsulated from one's present transactions with the world—how do they fit into the picture?

This paper attempts to approach some of these difficult questions indirectly, via an analysis of the principles by which cognition might have evolved. This broadly *telenomic* strategy—whereby cognitive processes are understood in terms of their fitness-enhancing properties—draws inspiration from Peter Godfrey-Smith's (1996) *environmental complexity thesis*. On this view, cognition evolved to coordinate organismic behaviour with certain complex (i.e. heterogeneous or variable) properties of the eco-niche. Thus construed, cognition functions to generate flexible patterns of behaviour in response to fluctuating environmental conditions.

We shall not dwell on the details of the environmental complexity thesis here. What interests us, rather, is how the general shape of Godfrey-Smith's explanatory framework—taken in conjunction with more recent advances in theoretical biology, computational neuroscience, and related disciplines—can inform contemporary philosophical debates about the nature of (biological) cognition. Drawing on insights afforded by these fields, we interpret complexity in terms of *uncertainty*, and suggest that distinctive profiles of adaptive plasticity emerge as the capacity to represent and anticipate various sources of uncertainty becomes increasingly more sophisticated. This analysis suggests behavioural flexibility per se is not sufficient to determine the cognitive status of an adaptive organism. Rather, we propose a narrower conception of cognition as a process rooted in a particular kind of functional organisation; namely, one that affords the capacity to model and interrogate counterfactual possibilities.

This paper is structured as follows: “[Homeostasis and the free energy principle](#)” begins by considering the homeostatic challenges posed by uncertain environments. We approach this topic from the perspective of the *free energy principle* (Friston 2010), a formal account of the autopoietic processes by which biological systems organise and sustain themselves as adaptive agents. “[Beyond homeostasis: Allostatics and hierarchical generative models](#)” outlines how the theoretical resources of the free energy principle extend to predictive (i.e. allostatic) forms of biological regulation. We focus on two complementary formulations

of allostasis, highlighting how these hierarchical control schemes inform fundamental questions about learning, planning, and adaptive behaviour. “[Biological regulation in an uncertain world](#)” examines the relation between environmental and biological complexity via an analysis of generative models. We sketch out three scenarios designed to illustrate how different kinds of model architecture endow distinctive capacities for the representation and resolution of uncertainty. Finally, “[Two options for cognition](#)” elaborates some of the key implications of this analysis for the concept of biological cognition. We argue that cognition does not simply coincide with adaptive biological activity (allostatic or otherwise), but inheres rather in the agent’s capacity to disengage from the present and entertain counterfactual states of affairs.

Homeostasis and the free energy principle

The free energy principle provides a mathematical framework explaining how adaptive organisms come to exist, persist, and thrive—at least for a while—by resisting what Schrödinger described as “the natural tendency of things to go over into disorder” (1992, p. 68). In this section, we sketch a relatively non-technical overview of this perspective, and show how it relates to familiar notions of homeostasis and adaptive behaviour.¹

Life, formalised: thermodynamics, attracting sets, and (un)certainty

The free energy principle starts with the simple (but fundamental) premise that organisms must maintain the stability of their internal dynamics in order to survive (Bernard 1974; Cannon 1929; Friston 2012a). This is to say that living systems must act to preserve their structural and functional integrity in the face of environmental perturbation (cf. autopoiesis; Maturana and Varela 1980), thereby resisting the tendency to disorder, dispersal, or *thermodynamic entropy* alluded to by Schrödinger (Friston 2013; Nicolis and Prigogine 1977).² Reformulated in the language of statistical mechanics: Living systems *live* in virtue of their capacity to keep themselves within some (nonequilibrium) thermodynamic steady-state. In other words, they maintain invariant (steady-state) characteristics far from equilibrium—as open systems in exchange with their environment.³

It follows from this postulate that any entity *qua* adaptive biological system can be expected to frequent a relatively small number of *attracting states*; namely

¹ For broader philosophical discussion of these ideas in the context of predictive processing, see Clark (2016), Hohwy (2013) and Wiese and Metzinger (2017). For more technical explications of the free energy principle and its corollaries, see Bogacz (2017), Buckley et al. (2017) and Friston et al. (2017a).

² Technically, living systems appear to violate *fluctuation theorems* that generalise the second law of thermodynamics to nonequilibrium systems (Evans and Searles 1994, 2002; Seifert 2012).

³ See Linson et al. (2018), for a lucid explication of the deep continuities between thermodynamics and the free energy principle. For a more technical exposition, see Sengupta et al. (2013).

those which compose its *attracting set* (Friston 2012a, 2013). In dynamical systems theoretic terms, this set of states corresponds to a *random dynamical attractor*, the invariant set towards which the system inevitably evolves over time (Crauel and Flandoli 1994). The existence of this invariant set means that the probability of finding the system in any given state can be summarised by a distribution (technically, an *ergodic density*), which can be interpreted in terms of its *information-theoretic entropy* or *uncertainty* (Shannon 1948).

The upshot of this picture is that any biotic (random dynamical) system which endures over time must do so in virtue of maintaining a low-entropy distribution over its attracting set (Friston 2012a; Friston and Ao 2012). This is tantamount to saying there is a high degree of certainty concerning the state of the system at any given moment in its lifetime, and that such attracting states will correspond to the conditions of the organism's homeostatic integrity. Conversely, there is a low probability of finding the system occupying a state outside of its attracting set, since such states are incompatible with the system's (long-term) existence. It follows that the repertoire of attracting states in which the system is typically located is constitutive of that agent's *phenotype* (Friston et al. 2009, 2010a), insofar as the phenotype is simply a description of the organism's characteristic (i.e. typically-observed) states.

Surprise and free energy minimisation

According to this framework, then, homeostasis amounts to the task of keeping the organism within the bounds of its attracting set (or, equivalently, of maintaining a *low conditional entropy* over its internal states). How might biological agents realise this outcome?

To answer this question, we must invoke another information-theoretic term: *surprise* (Shannon 1948). Surprise (i.e. 'surprisal' or self-information) quantifies the improbability (i.e. negative log-probability) of some outcome. In the present context, the outcome in question refers to some sensory state induced in any part of the system receptive to perturbation. Obvious realisers of sensory states include the sensory epithelia (e.g., retinal photoreceptor cells), but also extend to ion channel receptors in cell membranes, photosensitive receptors in plants, and so on. These receptive surfaces can be construed as states embedded within a (statistical) boundary or interface (technically, a *Markov blanket*; Pearl 1988) separating (i.e. 'shielding' or 'screening-off') system-internal from system-external conditions (see Friston 2013; Friston and Ao 2012; Hohwy 2017a).⁴

Importantly, the quantity of surprise associated with any given sensory state is not absolute, but depends rather on the kind of system the organism embodies (i.e. its phenotype or internal configuration; Friston and Stephan 2007). The fish that finds itself on dry land (i.e. well beyond the bounds of its attracting set) experiences a high degree of surprise, and will perish unless something is done (quickly!) to

⁴ Note that complex organisms may be composed of multiple, hierarchically-nested Markov blankets (for recent discussion, see Allen and Friston 2018; Clark 2017; Kirchhoff et al. 2018; Palacios et al. 2020; Ramstead et al. 2018).

reinstate its usual milieu. Conversely, this very same state will elicit relatively little surprise in land-dwelling creatures. It turns out that *minimising* or *suppressing* the surprise evoked by sensory states—that is, by avoiding surprising states and favouring unsurprising ones—the agent will tend to keep the (conditional) entropy of its states low, since entropy (almost certainly) converges with the long-term time average of surprise (Birkhoff 1931; Friston and Ao 2012).

In other words, by avoiding surprising interactions with their environment, biological systems keep themselves within the neighbourhood of attracting states that are conducive to their ongoing existence. Indeed, as a random dynamical system that repeatedly revisits its attracting set over time, the agent thereby realises itself as its own random dynamical attractor—and by extension, its own ‘existence proof’ (Friston 2018; more on which shortly).

There is, however, an important complication to this story: Surprise is computationally intractable, since its direct evaluation would require the agent to possess exhaustive knowledge of the external dynamics responsible for its sensory experiences (Friston 2009). This is where the concept of *free energy minimisation* comes in.

Variational free energy is an information-theoretic quantity developed to finesse difficult integration problems in quantum statistical mechanics (Feynman 1972).⁵ In the present context, free energy serves as a proxy for the amount of surprise elicited by sensory inputs (Friston 2010, 2011). As free energy is a function of the agent’s sensory and internal states (i.e. two sources of information available to the agent), and can be minimised to form a tight (upper) bound on sensory surprise, free energy minimisation enables the agent to indirectly evaluate the surprise associated with its sensory states (Friston and Stephan 2007). Moreover, since the agent is also capable of evaluating how free energy is likely to *change* in response to state transitions (Friston et al. 2012d), it will appear to select (or ‘sample’) actions that reduce surprise (Friston et al. 2015b).⁶ The free energy principle thus implies that biological systems will tend to avoid (or suppress) surprising observations over the long-run, thereby restricting themselves within the neighbourhood of their invariant (attracting) set.

Naturally, this explanation raises yet further questions: How does the agent minimise free energy to a ‘tight bound’ on surprise? How can simple organisms ‘expect’ to occupy certain states, or be said to ‘prefer’ these states over others? In order to address such questions, we first need to elaborate a notion of the agent as a *generative model*.

⁵ Variational inference techniques are also widely used in machine learning to approximate density functions through optimisation (see Blei et al. 2017).

⁶ Of course, just because a system can be *described* as behaving in a way that minimises variational free energy (maximises Bayesian model evidence, approximates Bayesian inference, etc.) does not guarantee that it *actually* implements any such computation. The extent to which the free energy principle should be construed as a useful heuristic for describing and predicting adaptive behaviour (a kind of *intentional stance*; Dennett 1987), versus a more substantive ontological claim, remains an open question. That said, recent progress has been made towards casting the free energy principle as a process theory of considerable explanatory ambition (Friston et al. 2017a).

Existence implies inference: agents as generative, self-evidencing models

According to the free energy principle, adaptive biological agents embody a probabilistic, generative model of their environment (Calvo and Friston 2017; Friston 2008, 2011, 2012a; Kirchhoff et al. 2018; Ramstead et al. 2018). As we shall see, this is a rather bold claim that moves us far beyond conventional accounts of homeostatic regulation⁷ and their reformulation in the language of statistical mechanics and dynamical systems theory.

Roughly, the system's form and internal configuration are said to parameterise a probabilistic mapping between the agent's sensory states and the external (hidden) causes of such states. This is to say that organisms interact with their eco-niche in ways that distill and recapitulate its causal structure, meaning that biological agents constitute (embody) a statistical model encoding conditional expectations about environmental dynamics (Allen and Friston 2018; Friston 2011; Kirchhoff et al. 2018).⁸ Indeed, according to the free energy principle, the very existence of the organism over time implies that it must optimise a generative model of the external causes of its sensory flows. This follows from the observation that optimising a model of the hidden dynamics impinging on one's sensory surfaces will give rise to (free-energy minimising) exchanges with the environment, which manifest as adaptive responses to evolving external conditions (Friston et al. 2006; Friston and Stephan 2007).

Under this account, then, even such simple biological agents as unicellular organisms will 'expect' (abstractly and nonconsciously) to find themselves in certain (unsurprising) states, according to the model they embody. Moreover, such agents will strive to sample (i.e. bring about) those attracting, free energy minimising states they expect to occupy—or risk perishing (Friston et al. 2006; Friston and Stephan 2007).

In Bayesian terms, this activity of expectation-fulfilment (or maximisation)—where expectations correspond to prior probability distributions parameterised by the agent's internal states—is tantamount to maximising the evidence for the agent's model (and by extension, their own existence; Friston 2010, 2013), a process known as *self-evidencing* (Hohwy 2016). Hence, under the free energy principle, adaptive biological systems conserve their own integrity through free energy minimising interactions which, over the long-term time average, minimise entropy (i.e. resolve uncertainty) and maximise self-evidence.⁹ The process by which they accomplish this feat is *active inference*.

⁷ Note that we interpret the notion of regulation rather broadly here. For philosophical arguments distinguishing regulation from related concepts such as feedback control and homeostasis, see Bich et al. (2016). On this view, regulatory control consists in a special kind of functional organisation characterised in terms of *second-order control*. This formulation seems broadly in line with our understanding of allostasis (see “[Beyond homeostasis: Allostasis and hierarchical generative models](#)”).

⁸ Note that the organism's morphology and internal organisation impose constraints on the way it models and represents environmental dynamics (e.g., Parr and Friston 2018a)—a point we shall elaborate in “[Biological regulation in an uncertain world](#)”.

⁹ See Parr and Friston (2018b) for a mathematical explanation of the (bound) relationship between variational free energy and model evidence.

Active inference: closing the perception–action loop

The scheme outlined above implies that biological agents conserve their morphology and internal dynamics (and in turn, the generative model these characteristics embody) by acting to offset the dispersive effects of random environmental fluctuations. But why should the agent sustain its model through such adaptive exchanges, rather than allowing its model to change in line with evolving environmental dynamics? As it turns out, the free energy principle supports both of these possibilities: agent and environment are locked in a perpetual cycle of reciprocal influence. This dialectical interplay, which emphasises the inherent *circular causality* at the heart of adaptive behaviour, is formalised under the active inference process theory (Friston et al. 2017a).

Active inference comprises two basic processes that play out at the agent–environment interface: perception and action.¹⁰ Here, perception is construed as the process of changing (i.e. ‘updating’) one’s internal states in response to external perturbations, and over longer timescales corresponds to learning (i.e. Bayesian updating of time-invariant model parameters; Fitzgerald et al. 2015; Friston et al. 2016, 2017a).¹¹ In other words, perceptual (state) inference describes how the agent updates its representation of environmental dynamics to resolve uncertainty about the hidden causes of its sensory fluctuations. A prevalent neurocomputational implementation of this scheme is *predictive coding* (Elias 1955; Lee and Mumford 2003; Rao and Ballard 1999; Srinivasan et al. 1982; Huang and Rao 2011; Spratling 2017; for some variational free energy treatments, see Barrett and Simmons 2015; Bastos et al. 2012; Friston and Kiebel 2009; Pezzulo 2014; Seth et al. 2012; Shipp et al. 2013; Shipp 2016).

Action, on the other hand, involves the activation of effector mechanisms (e.g., motor reflexes, cell migration; Friston et al. 2015a) in order to bring about new sensory states (Adams et al. 2013; Friston et al. 2010a). Different states can be sampled either through actions that directly intervene on the environment (e.g., turning off a bright light), or alter the relationship between the agent’s sensory surfaces and external states (e.g., turning away from a bright light). In either case, free energy is affected by the sensory consequences of the agent’s actions, where expectations

¹⁰ While active inference is sometimes narrowly construed as the active or behavioural component of the perception–action loop, the term was originally introduced to characterise the reciprocal interplay between perception and action (e.g., Friston et al. 2009, p. 4). This broader interpretation emphasises the deep continuity of the (Bayesian inferential) processes underwriting perception, learning, planning, and action under the free energy principle (Friston et al. 2017a).

¹¹ This general understanding of perception need not entail the conscious experience of sensations, just as learning can occur through entirely unconscious—and even artificial—mechanisms. Rather, what is at stake here is the statistical notion of *Bayesian belief*, where probability distributions encode the conditional probability that sensory observation *Y* was caused by hidden state *X*.

about the modifiability of sensory flows are conditioned on a model of hidden states and their time-evolving trajectories (Friston and Ao 2012).¹² Active inference thus recalls the cybernetic adage that organisms “control what they *sense*, not... what they *do*” (Powers 1973, p. 355, emphasis in original).

Although we shall have more to say about the role of action under active inference in later sections, these cursory remarks are sufficient to motivate the basic claim that adaptive agents recruit effector systems in order to propel themselves towards the sensory states they expect to inhabit.

Superficially at least, the inferential dynamics underwriting perception and action seem to pull in opposing directions (i.e. *change the model to reflect the world* vs. *change the world to reflect the model*). Under the active inference scheme, however, these two processes are complementary and deeply interwoven. This is because perception can only minimise free energy (or, under certain simplifying assumptions, *prediction error*; Friston 2009; Friston et al. 2007) to a tight (upper) bound on surprise, whereas action suppresses surprise by invoking new sensory states that conform to (expectations prescribed by) the agent’s phenotype. Consequently, perception serves to optimise the agent’s model of environmental conditions, such that the agent has adequate information to choose actions that engender low sensory entropy (Friston et al. 2010a).¹³

Although perceptual inference might seem to imply that agents ought to adapt their internal organisation to reflect environmental fluctuations as accurately as possible, unrestricted acquiescence to such dynamics would result in a precarious (and in many cases, rather brief) existence. Rather, the exigencies of homeostatic control dictate that biological systems preserve the *conditional independence* of their internal and external states (Ramstead et al. 2018). This is to say that the biological agent must maintain a boundary (i.e. Markov blanket) that separates (and insulates) its internal dynamics from external conditions.¹⁴ Consequently, the free energy minimising agent must exploit inferences about the state of the world beyond its Markov blanket in order to act in ways that keep it within the neighbourhood of its attracting states (Friston 2013).

The agent’s capacity to maintain the integrity of its Markov blanket is aided by prior beliefs about the sorts of conditions it expects to encounter. Many such expectations are directly functional to homeostasis (Pezzulo et al. 2015), having been

¹² Technically, actions are physical, real-world states that are not represented within the agent’s generative model (Attias 2003). Rather, the agent infers (fictive) ‘control’ states that explain the (sensory) consequences of its actions (Friston et al. 2012a, d). Action selection (or decision-making) thus amounts to the optimisation of posterior beliefs about the control states that determine hidden state transitions (Friston et al. 2013, 2015b).

¹³ Although one might be tempted to subordinate perceptual inference to free energy minimising action, we interpret perception and action as mutually dependent moments within a unified dynamical loop (cf. the perception–action cycle; Fuster 2001, 2004). Ultimately, *both* modes of active inference are in the service of uncertainty reduction: Percepts without actions are idle; actions without percepts are blind.

¹⁴ Formally speaking, the sensory and active states that compose the Markov blanket render the probability distributions over internal and external states statistically independent of one another (see Pearl 1988). In other words, internal and external states provide no additional information about one another once the Markov blanket’s active and sensory states are known.

shaped and refined through generations of natural selection (Allen and Friston 2018; de Vries and Friston 2017; Friston 2010). Pushing this logic one step further, we can say that the agent embodies a deeply-engrained expectation to survive (i.e. to remain within the confines of its attracting set—and thus to maintain its homeostatic integrity over time); this is simply the expectation to minimise average surprise over the long-run (Allen and Tsakiris 2018; Seth 2015). This remark highlights the point that not all beliefs are equally amenable to model updating. Rather, certain strongly-held or *high-precision* beliefs (e.g., those pertaining to homeostatic stability) will be stubbornly defended through actions that seek to substitute conflicting sensory evidence with input that conforms more closely to prior expectations (Yon et al. 2019).

In sum, perception and action work in concert to achieve free energy minimisation, ensuring that the biological system maintains itself in an invariant relationship with its environment over time. Critically, this formulation explains how apparently teleological or purposive behaviours emerge as a consequence of free energy minimising sensory sampling, without resorting to additional concepts such as ‘value’ or ‘reward’ (Friston et al. 2009, 2010a). Rather, value and reward simply fall out of the active inference process, as what is inherently valuable or rewarding for any particular organism is prescribed by the attracting states that compose its phenotype (i.e. those states the agent expects itself to occupy; Friston and Ao 2012). Simply put, unsurprising (i.e. expected) states are valuable; hence, minimising free energy corresponds to maximising value (Friston et al. 2012a).¹⁵

Beyond homeostasis: allostasis and hierarchical generative models

The free energy principle is founded on the premise that biological systems act to maintain their homeostatic equilibrium in the face of random environmental perturbations. Until recently, however, the question of how adaptive organisms secure their homeostatic integrity had attracted relatively little theoretical attention from within this perspective. A growing number of researchers are now leveraging predictive coding and active inference to explain how complex nervous systems monitor internal bodily states (i.e. perceptual inference in the *interoceptive* domain) and regulate physiological conditions (Allen et al. 2019; Barrett and Simmons 2015; Iodice et al. 2019; Seth 2013; Pezzulo 2014; for recent reviews, see Khalsa et al. 2018; Owens et al. 2018; Quadt et al. 2018).

An important conceptual development within this line of work was the move beyond traditional notions of homeostatic stability to more modern accounts of *allostatic* variability. The concept of allostasis (“stability through change”) was first introduced by Sterling and Eyer (1988), who criticised conventional homeostatic control theory as overly restrictive and reactive in character.¹⁶ By contrast, allostasis

¹⁵ Note that value here is not equivalent to expected utility, but rather a composite of utility (*extrinsic value*) and information gain (*epistemic value*; see Friston et al. 2015b; Schwartenbeck et al. 2015).

¹⁶ Although we focus here on allostasis, numerous other concepts emphasising the dynamic nature of biological regulation have been proposed in an effort to extend (or transcend) classical notions of homeostatic setpoint control (see for e.g., Bauman 2000; Berntson and Cacioppo 2000 and references therein).

was intended to replace setpoint defence with a more flexible scheme of parameter variation, and to supersede local feedback loops with centrally co-ordinated feed-forward mechanisms (e.g., *central command*; Dampney 2016; Goodwin et al. 1972; Krogh and Lindhard 1913). Allostasis was thus posited to account for a wide variety of anticipatory physiological activity that resisted explanation in terms of closed-loop control.

Despite controversy over the theoretical merits and conceptual scope of allostasis (see Corcoran and Hohwy 2018, for a recent overview), there is ample evidence that biological regulation consists in both anticipatory and reactive modes of compensation (see for e.g., Burdakov 2019; Ramsay and Woods 2016; Schulkin and Sterling 2019).¹⁷ These complementary mechanisms are easily accommodated within the active inference framework, mapping neatly onto the hierarchically-stratified models posited under the free energy principle (Friston 2008). Moreover, we believe that mature versions of allostatic theory are enriched and invigorated by active inference, insofar as the latter furnishes precisely the kind of inferential machinery required to underwrite effective forms of prospective control across various timescales (Corcoran and Hohwy 2018; Kiebel et al. 2008; Friston et al. 2017d; Pezzulo et al. 2018).

The remainder of this section briefly outlines two recent attempts to integrate homeostatic and allostatic mechanisms within the broader scheme of active inference. Although these perspectives assume a rather complex, neurally-implemented control architecture, we shall argue in “[Biological regulation in an uncertain world](#)” that the basic principles underwriting such schemes can be generalised to much simpler biological systems with relative ease.

Allostasis under active inference

Stephan and colleagues (Stephan et al. 2016, see also Petzschner et al. 2017) developed an active inference-based account of allostasis that maps interoception and physiological regulation onto a three-layer neural hierarchy. At the lowest level of this hierarchy are *homeostatic reflex arcs*, which operate much like classical feedback loops (i.e. deviation of an essential variable beyond certain limits elicits an error signal, which in turn triggers a countervailing effector response; see Ashby 1956, Ch. 12; Wiener 1961, Ch. 4). Critically, however, the range of states an essential variable may occupy is prescribed by intermediate-level *allostatic circuits*. This formulation thus recasts essential variable setpoints as (probabilistic) prior expectations (or equivalently, top-down model-based predictions) about the likely states of interoceptors (cf. Penny and Stephan 2014), with deviations from expected states provoking interoceptive prediction error.¹⁸

¹⁷ Indeed, evidence of anticipatory physiological regulation antedates Walter B. Cannon’s influential work—Ivan Pavlov’s (1902) Nobel prize-winning research on the digestive system demonstrated that gastric and pancreatic enzymes are secreted *before* nutrient ingestion (see Smith 2000; Teff 2011).

¹⁸ This formulation is congruent with contemporary efforts to finesse traditional notions of setpoint rigidity with more dynamic accounts of homeostatic control (e.g., Cabanac 2006; Ramsay and Woods 2014; cf. Ashby 1940). It also seems more felicitous to Cannon’s original conception of homeostatic control (see for e.g., Cannon 1939, p. 39).

Two important features of this account are that (1) prior expectations about essential variables encode a distribution over states (rather than a singular ideal reference value), and that (2) the sufficient statistics which specify this distribution—its mean and precision (inverse variance)—are free to vary (cf. Ainley et al. 2016). On this view, such classic allostatic phenomena as diurnal patterns of body temperature (Kräuchi and Wirz-Justice 1994) and blood pressure variation (Degaute et al. 1991) emerge as a consequence of the cyclical modulation of the priors over these physiological states (cf. Sterling 2004, 2012). Likewise, phasic increases or decreases in the stability of such variables correspond to periodic shifts between more- or less-precise distributions, respectively.¹⁹

Subordinating homeostatic reflex arcs to allostatic circuits transforms the traditional conception of physiological control as setpoint defence into a far more dynamic and context-sensitive process. Access to perceptual and cognitive representations (e.g., via the anterior insular and cingulate cortices; Barrett and Simmons 2015; Craig 2009; Gu et al. 2013; Menon and Uddin 2010; Paulus and Stein 2006) enables allostatic circuitry to harness multiple streams of information such that homeostatic parameters may be deftly altered in preparation for expected environmental changes (Ginty et al. 2017; Peters et al. 2017). Not only does this arrangement enable the system to anticipate periodic nonstationarities in essential variable dynamics (such as the circadian oscillations in body temperature and blood pressure mentioned above), it also confers potentially vital adaptive advantages under unexpected and uncertain conditions.

As a brief illustration, consider the case of an animal that detects the presence of a nearby predator. Registering its perilous situation, the animal's brain triggers a cascade of autonomic activity—the 'fight-or-flight' response famously characterised by Cannon (1914, 1915). On Stephan and colleagues' (2016) account, these rapid physiological alterations are mediated via the allostatic enslavement of homeostatic reflex loops. This generative model-based scheme explains why physiological parameters should change so dramatically in the *absence* of any immediate homeostatic disturbance: Predictions (or 'forecasts'; Petzschner et al. 2017) about the likely evolution of external conditions mandate the adoption of atypical, metabolically expensive states in preparation for evasive action (cf. Requin et al. 1991).

Notice that the physiological states realised via allostatic modulation of homeostatic loops might themselves constitute surprising departures from the organism's typically-expected states. Since these deviations cannot be resolved locally on account of the higher-order imperative to mobilise metabolic resources for impending action, interoceptive prediction error propagates up the neural hierarchy, possibly manifesting as the suite of sensations associated with acute stress (Peters et al. 2017). Such prediction error is tolerated to the extent that these emergency measures are expected to expedite a more hospitable environment (namely, one in which there is no immediate threat of predation). In other words, allostatic regimes of

¹⁹ Note that priors over certain physiological variables (e.g., core temperature, blood pH) are likely to be held with greater precision—and thus restricted to a narrower range of attracting states—than others (e.g., blood pressure, heart rate; see Allen and Tsakiris 2018; Seth and Friston 2016; Yon et al. 2019).

interoceptive active inference are functional to the agent's deeply-held expectation to survive, insofar as they serve to minimise uncertainty and maximise self-evidence *over the long-run*.²⁰

Stephan and colleagues (2016) crown their hierarchical framework with a *meta-cognitive* layer that monitors the efficacy of one's control systems. This processing level is posited to explain the emergence of higher-order beliefs about one's ability to adaptively respond to homeostatic perturbation. Persistent failure to suppress interoceptive surprise—either as a consequence of harbouring inaccurate allostatic expectations, or one's inability to realise free energy minimising actions—results in a state of *dyshomeostasis* (cf. *allostatic load*; McEwen and Stellar 1993; Peters et al. 2017), the experience of which may erode confidence in one's capacity for self-regulation. Stephan and colleagues (2016) speculate that the affective and intentional states engendered by chronic dyshomeostasis contribute to the development of major depressive disorder (cf. Badcock et al. 2017; Barrett et al. 2016; Seth and Friston 2016). Although such psychopathological implications are beyond the scope of this paper, the basic idea that the brain's homeostatic/allostatic architecture is reciprocally coupled with higher-order inferential processing will be explored further in “[Biological regulation in an uncertain world](#)”.

In sum, the hierarchical regulatory scheme proposed by Stephan and colleagues (2016) provides a promising formal description of the inferential loops underwriting both reactive (homeostatic) and prospective (allostatic) modes of biological regulation, and their interaction with higher-order beliefs. This framework accommodates a rich variety of allostatic phenomena spanning multiple timescales; ranging from deeply-entrenched, slowly-unfolding regularities (e.g., circadian and circannual rhythms) to highly unpredictable, transient events (e.g., predator–prey encounters), and everything in between (e.g., meal consumption; Morville et al. 2018; Teff 2011).

Broadening the inferential horizon: preferences, policies, and plans

A second, complementary perspective focuses on the ways organisms can develop complex behavioural repertoires that optimise physiological regulation in an anticipatory manner (e.g., buying food and preparing a meal before one is hungry).

Active inference agents can acquire such skills by leveraging information about evolving state transitions, or *policies*. Policies are (beliefs about) sequences of actions (or more precisely, *control states*; see Footnote 12) required to minimise free energy in the future, thereby realising some preferred (i.e. expected, self-evidencing, and thus *valuable*) outcome (Attias 2003; Friston et al. 2012a, 2013; Pezzulo et al. 2018). In active inference, policies are explicitly evaluated (and therefore selected)

²⁰ One might protest that all we have done here is pivot from one sort of reactive homeostatic mechanism to another; albeit, one involving responses to an external (rather than internal) threat. Nevertheless, we consider this simple scenario as exemplary of the fundamental principle of allostatic regulation; namely, the modulation of physiological states in anticipation of future conditions, and in the absence of any immediate homeostatic perturbation. This example can easily be extended to capture a rich assortment of allostatic dynamics that play out across increasing levels of abstraction and spatiotemporal scale.

depending on their *expected free energy*, i.e. the amount of free energy they are expected to minimise in the future. It is important not to conflate this notion of expected free energy with that of *variational* free energy (as introduced in “[Surprise and free energy minimisation](#)”). The former only arises during policy evaluation and uses *expectations* about future states of affairs that may arise from selecting a particular policy; whereas the latter uses (available) information about past and present states of affairs.

Policy selection is important for allostatic control, because by explicitly considering future states of affairs in addition to one’s immediate needs, agents can (learn how to) engage in relatively complex courses of action that minimise more free energy over the long-run. Consider for instance the decision to purchase ingredients from a local supermarket and return home to cook a meal, versus ordering a meal from a neighbouring fast food restaurant. In both cases, the underlying homeostatic motivation driving behaviour (i.e. increasing prediction error manifesting as intensifying hunger) is identical; the interesting question is why one does not always opt for the policy that is most likely to resolve prediction error (hunger) most rapidly. Selecting the *Cook* policy, which postpones the resolution of interoceptive prediction errors (and thus engenders greater free energy in the short-term), might appear on first blush to contradict the free energy principle. Such choices can however be explained by recourse to the agent’s superordinate expectation to minimise expected free energy over longer timescales (e.g., prior beliefs about the health, financial, and/or social benefits associated with domestic meal preparation; cf. Friston et al. 2015b; Pezzulo 2017; Pezzulo et al. 2018).²¹

Pezzulo and colleagues (2015) offer an account of allostasis that seeks to explain the gamut of behavioural control schemes acquired via associative learning from a unified active inference perspective.²² Specifically, this account grounds the emergence of progressively more flexible and sophisticated patterns of adaptive behaviour on evolutionarily primitive control architectures (e.g., low-level circuitry akin to Stephan and colleagues’ (2016) homeostatic reflex arc). From a broader ethological perspective, this scheme implies a deep continuity between the homeostatic loops underpinning simple, stereotypical response behaviour on the one hand, and the complex processes supporting goal-directed decision-making and planning on the other.

According to this view, all associative learning-based control schemes fall out of the same uncertainty-reducing dynamics prescribed by the free energy principle. What distinguishes these schemes under the active inference framework is their place in the model hierarchy: While rudimentary adaptive behaviours (e.g., approach/avoidance reflexes) are availed by ‘shallow’ architectures, more sophisticated modes of control require greater degrees of hierarchical depth. Goal-directed actions require generative models that are capable of representing the prospective

²¹ Note that the appeal to expected free energy was also implicit in the predator example of the previous section, insofar as transient increases in homeostatic prediction error were tolerated in order to avoid a much more surprising fate—being eaten!

²² See Moore (2004) for a thoroughgoing review of such associative learning mechanisms.

evolution of hidden states over sufficiently long intervals (cf. Botvinick and Toussaint 2012; Penny et al. 2013; Solway and Botvinick 2012), while simultaneously predicting how these projected trajectories are likely to impact upon the internal states of the organism (cf. Keramati and Gutkin 2014). On this account, activity at higher (or deeper) hierarchical layers (e.g., prefrontal cortical networks) contextualises that of more primitive control schemes operating at lower levels of the hierarchy (see also Pezzulo and Cisek 2016; Pezzulo et al. 2018). This means that higher-level inferences about distal or remote states (and the policies most likely to realise them) inform lower-level mechanisms governing action over shorter timescales (see also Attias 2003; Badre 2008; Friston et al. 2016; Kaplan and Friston 2018; Pezzulo et al. 2018).

A distinctive feature of Pezzulo and colleagues' (2015) scheme is the crucial role played by the (cross- or multimodal) integration of interoceptive, proprioceptive, and exteroceptive information over time. This is required if one wants to translate inferences on time-varying internal states (e.g., declining blood glucose concentration) into complex behavioural strategies (e.g., preparing a meal) that anticipate or prevent homeostatic disturbance. This is to say that the emergence of nervous systems which enable their owners to envisage and pursue certain future states at the expense of others depends upon the (allostatic) capacity to track and anticipate co-evolving internal/sensory and external/active state trajectories.²³ In short, Pezzulo and colleagues (2015) posit that hierarchical generative models harness prior experience to map sensorimotor events to interoceptive fluctuations. This mapping enables the agent to learn how their interoceptive/affective states are likely to change both endogenously (e.g., *I am likely to become irritable if I forgo my morning coffee*), and in the context of external conditions (e.g., *I am likely to dehydrate if I exercise in this heat without consuming fluids*).²⁴

With this (hierarchical) inferential architecture in place, it is relatively easy to see how allostatic policies may take root. As alluded to above, interoceptive/homeostatic dynamics often exhibit (quasi)periodic cycles, thus facilitating the modelling and prediction of time-evolving changes in internal sensory states. Given a model of how interoceptive states typically oscillate, the agent learns how particular external perturbations (including those caused by its own actions) modulate this trajectory (cf. Allen and Tsakiris 2018). As the agent accrues experience, it progressively refines

²³ More precisely, this capacity depends on the ability to infer the expected free energy of the outcomes associated with various potential state trajectories, as well as the expected likelihood of outcomes under each policy (see Friston et al. 2017a, c; Parr and Friston 2017, 2018b). We have suggested such inferential processes might be facilitated by the co-ordination of exteroceptive sampling and motor activity with periodic regimes of autonomic/interoceptive active inference (Corcoran et al. 2018).

²⁴ We emphasise again that the conscious, reflective character of these intuitive examples should not detract from the idea that the *possibility* of such experiences is underwritten by more basic, unconscious allostatic mechanisms. For example, the growth onset of a horse's winter coat is not assumed to represent a strategic decision on the part of the horse, but rather a physiological response to seasonal changes in photoperiod. Similarly, a rabbit might schedule her foraging bouts to balance energy gain against predation risk, even though she might not be capable of representing and evaluating these concerns explicitly (this trade-off may, for instance, be implicitly encoded within the animal's circadian rhythm—see “Model 2: Hierarchical active inference”).

its model of the contingent relations that obtain between sensorimotor occurrences and physiological fluctuations, engendering the ability to extrapolate from sensations experienced in the past and present to those expected in the future (Friston et al. 2017a). This capacity is not only crucial for finessing the fundamental control problems posed by homeostasis (i.e. inferring the optimal policy for securing future survival and reproductive success), but also for its vital contribution in establishing the agent's understanding of itself *qua* autonomous agent (cf. Fotopoulou and Tsakiris 2017; Friston 2017). It is a relatively small step from here to the emergence of goal-directed behaviours that are ostensibly independent of (i.e. detached or decoupled from) current stimuli, hence permitting anticipatory forms of biological regulation (e.g., purchasing food when one is *not* hungry; see Pezzulo and Castelfranchi 2009; Pezzulo 2017).

Interim summary

In this section, we have presented two closely-related computational perspectives on biological regulation that cast homeostasis and allostasis within the broader scheme of active inference. We believe these accounts can be productively synthesised into a comprehensive framework that explains the emergence of increasingly versatile, context-sensitive, and temporally-extended forms of allostatic regulation. This framework provides a formal account of biological regulation that eschews the conceptual limitations of setpoint invariance (see Cabanac 2006; Ramsay and Woods 2014), unifies habitual ('model-free') and goal-directed ('model-based') behaviour (Dolan and Dayan 2013) under a single hierarchical architecture (see Fitzgerald et al. 2014; Pezzulo et al. 2016), and converges with neurophysiologically-informed perspectives on mind–body integration (e.g., Critchley and Harrison 2013; Smith et al. 2017). We have also introduced the important notion of policy selection, which explains how adaptive behaviour emerges through (active) inference of beliefs about the future (cf. 'planning as inference'; Attias 2003; Botvinick and Toussaint 2012; Solway and Botvinick 2012).

From a broader perspective, the capacity of higher model levels to track the evolution of increasingly distal, temporally-extended, and abstract hidden dynamics, and to infer the likely consequences of such dynamics for the agent's own integrity and wellbeing, provides a compelling explanation of how allostatic control schemes could have established themselves over ontogenetic and phylogenetic timescales. Not only does this perspective provide a principled account of how allostatic mechanisms should 'know' when to initiate adaptive compensations in the absence of physiological disturbance (i.e. how the body 'acquires its wisdom'; Dworkin 1993), the embedding of such processes within an overarching hierarchical model also explains how agents are able to effectively arbitrate and trade-off multiple competing demands (a core feature of many allostatic frameworks; e.g., Sanchez-Fibla

et al. 2010; Sterling 2012; Schulkin and Sterling 2019; Verschure et al. 2014).²⁵ In the next section of this paper, we consider *why* such allostatic regimes should have evolved.

Biological regulation in an uncertain world

We have argued that adaptive biological activity is underwritten by active inference, where more sophisticated (predictive or prospective) forms of biological regulation are supported by increasingly more sophisticated generative models that extract and exploit long-term, patterned regularities in internal and external conditions. In this section, we take a closer look at how the functional organisation of the inferential architecture constrains the organism's capacity to represent time-evolving state trajectories, and the impact this has upon its ability to deal with uncertainty.

Our analysis draws inspiration from Peter Godfrey-Smith's influential *environmental complexity thesis* (1996), which casts cognition as an adaptation to certain complex (i.e. heterogeneous or variable) properties of the organism's eco-niche. On this view, cognition evolved to mitigate or 'neutralise' environmental complexity by means of *behavioural complexity*—"the ability to do a lot of different things, in different conditions" (Godfrey-Smith 1996, p. 26).²⁶

The concept of complexity at the core of Godfrey-Smith's analysis is deliberately broad and abstract. Environments may comprise manifold dimensions of complexity, many of which may be of no ecological relevance to their inhabitants. Patterns of variation only become biologically salient once the capacity to track and co-ordinate with them confers a selective advantage (i.e. when sensitivity to environmental variation helps the organism to solve problems—or exploit opportunities—that bear on its fitness; Godfrey-Smith 2002). Much like the notion of surprise (conditional entropy) introduced in "[Surprise and free energy minimisation](#)", then, the implications of environmental complexity for any given organism are determined by the latter's constitution and relation to its niche.

In what follows, we analyse the connection between environmental and behavioural complexity as mediated by increasingly elaborate schemes of active inference. Following Godfrey-Smith's observation that complexity can be cast as "disorder, in the sense of uncertainty" (1996, p. 24; see also pp. 153–154), we consider how the

²⁵ See Morville et al. (2018) for discussion of the nontrivial challenges posed by high-dimensional homeostatic needs in uncertain environments. The ability to reliably navigate such complex demands speaks also to the notion of *competence* in artificial intelligence research (see Miracchi 2019).

²⁶ This gloss on the environmental complexity thesis is reminiscent of W. Ross Ashby's *law of requisite variety* (1956, 1958; cf. Conant and Ashby 1970), and is clearly in line with recent neuroscientific interest in the brain's teleonomic function as a sophisticated biological regulator (for discussion, see Williams and Colling 2018). Although Godfrey-Smith (1996, pp. 76–79) briefly remarks upon the connection between cybernetic accounts of homeostatic control and cognitive function, he rejects their strong continuity on the grounds that cognition can sustain biological viability through actions that circumvent homeostatic mechanisms. We concur that non-trivial definitions of homeostasis and cognition invoke concepts that are distinct from one another, and argue below that this distinction can be cashed out in terms of their constitutive inferential architectures.

exigencies of biological regulation under conditions of uncertainty may have promoted the evolution of increasingly more complex inferential architectures, and how such architectures enable organisms to navigate complex environments with increasing adroitness.

To this end, we will consider three successive forms of generative model that may underwrite different sorts of creatures. First, we take a *simple* generative model—and implicit architecture for active inference—that may be suitable for explaining single-celled organisms that show elemental homeostasis and reflexive behaviour. We then consider *hierarchical* generative models that have parametric depth, in the sense that they afford inference at multiple timescales (where faster dynamics at lower levels are contextualised by slower dynamics at higher levels). This produces adaptive systems that evince a deep temporal structure in their exchange with the environment by simply minimising free energy. An illustrative example of this in the active inference literature is birdsong; namely, the generation and recognition of songs that have an elemental narrative with separation of temporal scales (Kiebel et al. 2008). We will use this hierarchical scheme to explain certain aspects of allostasis such as circadian regulation, which permits the agent to implicitly track and adapt metabolic operations to slow temporal dynamics (i.e. cycles of night and day).

The third kind of generative model supplements parametric depth with *temporal depth*, or the ability to engage in counterfactual active inference. It is important to note that agents that are endowed with parametrically (but not temporally) deep models are quite limited; they can infer and adapt to future circumstances, but cannot actively select which one to attend. For example, although birds can recognise particular songs of conspecifics, this form of perceptual inference does not entail actively attending to one bird or another. In other words, it does not entail a selection among ways in which to engage with the sensorium. To bring this kind of selection into the picture, one needs to evaluate the expected free energy following one or another action (e.g., attending to one bird or another). However, in order to evaluate expected free energy, one has to have a generative model of the future—that is, the consequences of action. This in turn calls for generative models with temporal or counterfactual depth that are necessary to evaluate the expected free energy of a given policy. It is this minimisation of expected free energy—that converts sentient systems into agents that reflect and plan, in the sense of entertaining the counterfactual outcomes of their actions—that we associate with cognition.

Model 1: Minimal active inference

First, let us consider a simple example of homeostatic conservation through a ‘minimal’ active inference architecture.²⁷ We model this ‘creature’ on simplified aspects of *Escherichia coli* (*E. coli*) bacteria to emphasise the generality of such schemes beyond neurally-implemented control systems.

²⁷ See Baltieri and Buckley (2017) and McGregor et al. (2015) for alternative formulations of ‘minimal’ active inference.

Our *E. coli*-like creature is a unicellular organism equipped with a cell membrane (i.e. a Markov blanket separating internal from external states), a metabolic pathway (i.e. an autopoietic network that harnesses thermodynamic flows to realise and replenish the organism's constitutive components), and a sensorimotor pathway; but at the outset nothing approximating a nervous system (actual *E. coli* is of course much more complicated than this). Cellular metabolism depends on the agent's ability to absorb sufficient amounts of nutrient (e.g., glucose) from its immediate environment. However, the distribution of nutrient varies across the environment, meaning the agent must seek out nutrient-rich patches in order to survive. Like real *E. coli*, our creature attempts to realise this goal by alternating between two chemotactic policies: *Run* (i.e. swim along the present course) versus *Tumble* (i.e. randomly reorient to a new course, commence swimming; see Fig. 1).

Our simplified *E. coli*-like creature embodies a model that encodes an expectation to inhabit a nutrient-rich milieu. Variation in the environment's chemical profile means that this expectation is not always satisfied—sometimes the agent finds itself in regions where chemical attractant is relatively scarce. Crucially, however, the organism can infer its progress along the nutrient gradient through periodic sampling of its chemosensory states, and acts on this information such that it tends to swim up the gradient over time.²⁸

This rudimentary sensorimotor control architecture affords the agent a very primitive picture of the world—one that picks out a single, salient dimension of environmental complexity (i.e. attractant rate of change). The capacity to estimate or infer this property implies a model that prescribes a fixed expectation about the kind of milieu the agent will inhabit, while also admitting some degree of uncertainty as to whether this expectation will be satisfied at any given moment. The task of the agent is to accumulate evidence in favour of its model by sampling from its policies in such a way that it ascends the nutrient gradient, thereby realising its expected sensory states (cf. Tschantz et al. 2019).

Although severely limited in terms of the perceptual or representational capacities at its disposal, this need not imply suboptimality per se. Consider the case in which various kinds of attractant are compatible with the organism's chemoreceptors. The agent cannot discriminate amongst these chemical substances; all it can do is infer the presence (or absence) of 'nutrient' at its various receptor sites. Assuming all forms of chemical attractant are equally nutritious (i.e. equally 'preferable' or 'valuable' given the agent's phenotype), this source of environmental heterogeneity

²⁸ In fact, real *E. coli* realise a similar 'adaptive gradient climbing' strategy by integrating chemosensory information about the ambient chemical environment over time, and modulating the probability of tumbling as a function of attractant rate of change (Berg and Brown 1972; Falke et al. 1997). More recent work has indicated that such chemotactic activity approximates optimal Kalman filtering (Andrews et al. 2006), where hidden states are estimated on the basis of prior and present observations weighted by their uncertainty (Kalman 1960; Kalman and Bucy 1961; see Grush 2004, for discussion). As Kalman filtering constitutes a special case of Bayesian filtering (one that is equivalent to predictive coding; Bastos et al. 2012; Friston et al. 2010b, 2018), chemotaxis can be cast as a gradient descent on variational free energy. Notice that our model is deliberately simpler than this scheme, since sensory prediction errors are not modulated by an uncertainty (precision) parameter.

turns out to be entirely irrelevant to the system's ongoing viability. Consequently, the extra structural and functional complexity required to distinguish these substances would afford the organism no adaptive benefit—on the contrary, the additional metabolic costs incurred by such apparatus might pose a hindrance.²⁹

Our *E. coli*-like creature thus trades in a rather coarse representational currency, thereby minimising the costs associated with unwarranted degrees of organisational complexity. This is an example of optimising the trade-off between model accuracy and complexity (Fitzgerald et al. 2014; Hobson and Friston 2012; Moran et al. 2014), where the simplest model to satisfactorily explain observed data (i.e. the presence/absence of nutrient) defeats more complex competitors (or on an evolutionary timescale, where natural selection favours the simplest model that suffices for survival and reproductive success; Campbell 2016; Friston 2018). This also explains why some creatures might have evolved *simpler* phenotypes from more complicated progenitors—natural selection 'rewards efficiency' over the long-run (McCoy 1977).

This caveat notwithstanding, there remain a great many aspects of the environment that the minimal active inference agent fails to model *despite* their potential bearing on its wellbeing. One such omission is the system's incapacity to represent the evolution of its states over multiple sensory samples. This limitation is significant, since it prevents the organism from discerning patterns of variation over time, which in turn renders it overly sensitive to minor fluctuations in prediction error. For instance, the organism might trigger its *Tumble* policy at the first sign of gradient descent, even though this decrement might stem from a trivial divergence in the quantity of attractant detected across sensory samples. Unable to contextualise incoming sensory information with respect to the broader trajectory of its sensory flows, the agent risks tumbling out of a nutrient-rich stream due to innocuous or transient instability of the gradient, or due to the random error introduced by inherently noisy signalling pathways.

Relatedly, the agent's inability to retain and integrate over past experiences precludes the construction of map-like representations of previously-explored territory. The organism thus loses valuable information about the various conditions encountered on previous foraging runs—information that a more sophisticated creature could potentially exploit in order to extrapolate the most promising prospects for future forays. Moreover, it also lacks the necessary model parameters to track various distal properties that modulate or covary with the distribution of attractant (e.g., weather conditions, conspecifics, etc.). The agent is thus unable to exploit the patterned regularities that obtain between proximal and distal hidden states, and that afford predictive cues about the likely consequences of pursuing a particular policy (cf. fish species whose swim policies are informed by predictions about distal

²⁹ The story changes if the organism's receptors are compatible with molecules it cannot metabolise, or that afford low nutritional value (assuming such molecules are prevalent enough to significantly interfere with chemotaxis). See Sterelny (2003, pp. 20–26) for discussion of the challenges posed by 'informationally translucent environments' that confront organisms with ambiguous (or misleading) cues. Environmental translucence calls for greater model complexity; e.g., the capacity to integrate information harvested across multiple sensory channels (cf. *robust tracking*; Sterelny 2003, pp. 27–29).

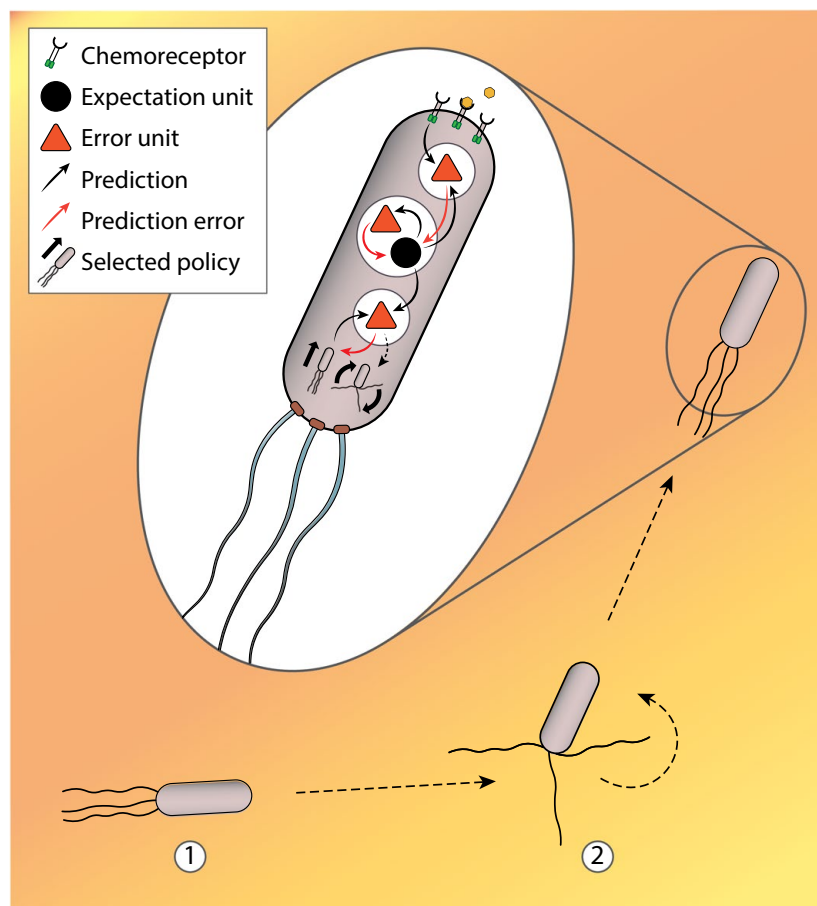


Fig. 1 A simple active inference model of bacterial chemotaxis. This figure depicts a simple active inference agent that must sample from its sensory states in order to infer the best course of (chemotactic) action. Since the organism expects its transmembrane chemoreceptors to be occupied by attractant molecules, absence of attractant at these sites evokes prediction error (red triangles). These signals are projected (e.g., via protein pathways; red arrows) to the agent's motor control network, where they are summed and compared to the expectation induced by the previous wave of sensory input (black circle). If the prediction error generated by current sensory input is reduced relative to that of the preceding cycle of perceptual inference, this constitutes evidence that the agent is ascending the nutrient gradient; i.e. evidence favouring the *Run* policy (1). Conversely, increased prediction error furnishes evidence of gradient descent, thus compelling the agent to sample from its *Tumble* policy (2). Here, policies are enacted via prediction errors that induce clockwise (*Tumble*) or anti-clockwise (*Run*) flagellar motion. Note that the organism's metabolic system has been omitted from this schematic. Figure reproduced from Corcoran et al. 2019 (CC BY 4.0)

feeding conditions and temperature gradients; Fernö et al. 1998; Neill 1979). Unable to 'see' beyond the present state of its sensory interface with the world, the organism has no option but to tumble randomly towards a new, unknown territory each time prediction error accrues.

In sum, the agent we have described here embodies a very simple active inference scheme; one which supports adaptive responses to an ecologically-relevant dimension of environmental complexity. While the agent does not always succeed in inferring the best chemotactic policy in a given situation, its strategy of alternating between active states in accordance with local nutrient conditions is cheap

and efficient, and tends to prevent it from drifting too far beyond its attracting set. But the severe epistemic constraints enforced by the agent's extremely narrow representational repertoire—both in the sense of its highly constricted spatiotemporal horizon, and in the poverty of its content—render this organism a creature of hazard. Unable to profit from past experience or future beliefs, it is locked in a perpetual present. This creature is thus thoroughly homeostatic in nature, activating its effector mechanisms whenever error signals indicate deviation beyond setpoint bounds.³⁰

Before moving onto our next model, let us briefly consider whether a creature could exist by simply maintaining its homeostatic stability in the absence of interoceptive modelling and action.³¹ When a creature of this sort encounters surprising deviations from its homeostatic expectations it only ever adjusts its internal states, never its active states. It may for instance change its metabolic rate (e.g., slow respiration, inhibit protein synthesis) in response to altered nutrient conditions, rather than acting on the environment in order to reinstate homeostatic equilibrium.³²

It is difficult to see how such a creature could actually exist in anything but a transitory, serendipitous manner. Changing its internal states in response to interoceptive prediction error is tantamount to yielding entirely to uncertainty. For example, as the nutrient gradient declines the organism's metabolic rate keeps decreasing, until it eventually starves to death—its states disperse throughout all possible states. An organism that fails to act upon its environment is ill-placed to avoid surprise and resist entropy. Only by happening to occupy a perfectly welcoming niche could it survive, but this is just to assume an environment devoid of uncertainty—not our world.³³

³⁰ Indeed, one might construe the minimal model as a simplified analogue of Ashby's (1960) 'Homeostat'.

³¹ See Godfrey-Smith (2016b) for a complementary discussion of this topic in relation to microbial proto-cognition and metabolic regulation.

³² One might call this entity a *Spencerian creature*; i.e. an organism that responds to environmental change through "the continuous adjustment of internal relations to external relations" (Spencer 1867, p. 82; see discussion in Godfrey-Smith 1996, pp. 70–71). From an active inference perspective, this creature is the embodiment of pure perception; i.e. an organism that reconfigures its internal states (updates its model) in accordance with external conditions, without ever seeking to alter such conditions (cf. Bruineberg et al. 2018; Corcoran 2019).

³³ One might play with the idea of entities that could exist like this quite happily once the ideal, invariant niche is discovered—perhaps deep within rocky crevices or underwater (one is reminded of the sea squirt that consumes its own brain after settling upon a permanent home, but the anecdote turns out to be an exaggeration; see Mackie and Burighel 2005). However, entities of this sort would surely fail to qualify as *adaptive* biological systems—at least insofar as the notion of adaptability implies some capacity to maintain one's viability in the face of time-varying environmental dynamics (cf. 'mere' vs. 'adaptive' active inference; Kirchhoff et al. 2018). Moreover, such entities would also fail to qualify as *agents* in any biologically relevant sense (see for e.g., Moreno and Etxeberria 2005).

Interestingly, this scenario is reminiscent of a common criticism levelled against the free energy principle: the so-called *dark-room problem* (Friston et al. 2012e). The thrust of this argument is that free energy minimisation should compel agents to seek out the least-surprising environments possible (e.g., a room devoid of stimulation) and stay there until perishing. Various rejoinders to this charge have been made (see for e.g., Clark 2018; Hohwy 2013; Schwartenbeck et al. 2013), including the observation that this strategy will inevitably lead to increasing free energy on account of accumulating interoceptive prediction error (Corcoran 2019; Pezzulo et al. 2015). More technically, "itinerant dynamics in the environment preclude simple solutions to avoiding surprise" (Friston et al. 2009, p. 2), where the environment referred to here includes the biophysical conditions that obtain *within* the organism, as well as without.

Model 2: Hierarchical active inference

Next, let us consider a more elaborate version of our creature, now equipped with a more sophisticated, *hierarchical* generative model of its environment—one which captures how environmental dynamics unfold over multiple timescales. Because higher levels of the generative model subtend increasingly broad temporal scales (Friston et al. 2017d; Kiebel et al. 2008), we shall see that this creature is capable of inferring the causes of slower fluctuations in the nutrient gradient. An implication of this arrangement is the emergence of parameters encoding higher-order expectations about the content and variability of sensory flows over time (cf. the fixed expectation of a high-nutrient state in Model 1).

In the interests of tractability, we limit ourselves to a fairly schematic illustration of hierarchical active inference in the context of circadian regulation. Circadian processes are near ubiquitous features of biological systems (even bacteria like *E. coli* show evidence of circadian rhythmicity; Wen et al. 2015), and provide a useful example of how internal dynamics can be harnessed to anticipate environmental variability.

Circadian clocks are endogenous, self-sustaining timing mechanisms that enable organisms to co-ordinate a host of metabolic processes over an approximately 24 h period (Bailey et al. 2014; Dyar et al. 2018). From an allostatic perspective, circadian oscillations furnish a temporal frame of reference enabling the organism to anticipate (and efficiently prepare for) patterned changes in ecologically-relevant variables (e.g., diurnal cycles of light and temperature variation).³⁴ We can incorporate a molecular clock within our active inference agent by installing oscillatory protein pathways within its metabolic network (Nakajima et al. 2005; Rust et al. 2007; Zwicker et al. 2010). With this timing mechanism in place, our creature may begin to track systematic variations in the temporal dynamics of its internal and sensory states.

Suppose our organism exists in a medium that becomes increasingly viscous as temperature declines overnight. The impact of these environmental fluctuations is two-fold: Colder ambient temperatures cool the organism, slowing its metabolic rate; greater viscosity increases the medium's resistance, making chemotaxis more energy-intensive. Initially, the agent might interpret unexpectedly high rates of energy expenditure as indicative of suboptimal chemotaxis, thus compelling it

Footnote 33 (continued)

This is to say that the attractors around which adaptive biological systems self-organise are inherently unstable—both *autopoietic* ('self-creating') and *autovitiating* ('self-destroying')—thus inducing itinerant trajectories (*heteroclinic cycles*) through state-space (Friston 2011, 2012b; Friston and Ao 2012; Friston et al. 2012c).

In other words, dark rooms may very well appeal to creatures like us (e.g., as homeostatic sleep pressure peaks towards the end of the day), but the value such environments afford will inevitably decay as alternative possibilities (e.g., leaving the room to find breakfast after a good night's sleep) become more salient and attractive (cf. *alliesthesia*, the modulation of affective and motivational states according to (time-evolving) physiological conditions; Berridge 2004; Cabanac 1971).

³⁴ Note that the allostatic treatment of circadian regulation may in principle be extended to periodic phenomena spanning shorter or longer timescales; e.g., ultradian and circannual rhythms.

to sample its *Tumble* policy more frequently in an effort to discover a nutrient-rich patch. Over time, however, the agent may come to associate a particular phase of its circadian cycle with higher average energy expenditure *irrespective* of policy selection. Our creature can capitalise on this information by scheduling its more expensive metabolic operations to coincide with warmer times of day, while restricting its nocturnal activity to a few essential chemical reactions. In other words, the agent can reorganise its behaviour (i.e. develop a rudimentary sleep/wake cycle) in order to improve its fit with its environment.³⁵

This scenario is indicative of how a relatively simple hierarchical agent may come to model time-varying hidden states in the distal environment. Like its minimal active inference counterpart, the hierarchical agent registers fluctuations in its sensory and internal states, and responds to them appropriately given its available policies. Unlike the minimal agent, however, these rapid fluctuations are themselves subject to second-order processing, in which successive sensory samples are integrated under a probabilistic representation of first-order variation (see Fig. 2). The ability to contextualise faster fluctuations in relation to the slower oscillatory dynamics of the circadian timekeeper enables the agent to infer that it is subject to periodic environmental perturbations, the origin of which can be parsimoniously ascribed to some unitary external process.³⁶ This example hints at a central tenet of the active inference scheme; namely, that the hierarchical organisation of the generative model implies a hierarchy of temporal scales, where causal dynamics subtending larger timeframes are encoded at higher levels of the model (Friston 2008; Friston et al. 2017d; Kiebel et al. 2008).

The hierarchical picture we have sketched here speaks to two complementary aspects of *representational detachment* (cf. Gärdenfors 1995; Pezzulo and Castelfranchi 2007; Pezzulo 2008) engendered by allostatic architectures. First, the separation of processing layers within the model hierarchy gives rise to a kind of temporal decoupling, in which higher layers construct extended representations of low-level sensory states. Although it might be tempting to think of these representations as aggregates of successive sensory samples, this does not do justice to the sophisticated nature of perception under active inference. Rather, higher layers of the hierarchy are perpetually engaged in modelling the evolution of the organism's sensory and internal states, and thus inferring the probable motion of the distal causes of its sensory flows. Consequently, higher-order representations 'reach out' beyond

³⁵ This scenario is not meant to imply that circadian rhythms are actually acquired in this fashion (although they are clearly susceptible to modulation through external cues). Rather, the idea we are trying to illustrate here is the way hierarchical architectures ground adaptive regulation over longer time-scales by dint of their capacity to capture recurrent, slowly evolving patterns of environmental variation.

³⁶ Notice that the agent forms a representation of a hidden cause corresponding to diurnal patterns of temperature variation *despite* its lack of exteroceptive sensitivity to such variables as temperature, viscosity, light, etc. Rather, it detects regular changes in its dynamics that cannot be ascribed to its own actions (which average out across the 24 h period), and infers some hidden external process as being responsible for these changes. It might not be right to say the agent represents ambient temperature per se, nor indeed the higher-order causes of the latter's oscillation (sun exposure, planetary rotation, etc.). Our agent lacks sufficient hierarchical depth to arrive at such conclusions, collapsing these fine-grained distinctions into a fairly 'flat', undifferentiated representation of diurnal variation.

the limits of each sensory moment, extrapolating forwards and backwards in time to synthesise an expanded temporal horizon (see Fig. 2a).

Second, there is a related sense in which higher-level processing within the hierarchy realises a more negative or reductive kind of detachment from low-level sensory input. Higher-level representations do not merely recapitulate (and predict) the bare contents of sensory experience, but seek instead to extract patterned continuities amidst the flux of sensory stimulation. This is to say that higher levels of the model attempt to carve out biologically-relevant signals within the agent's environment, while dampening or discarding the remaining content of sensory flows. This again speaks to the tension between model accuracy and complexity: Good models capture real patterns of environmental complexity, without being overly sensitive to the data at hand (and thus at risk of accruing prediction error over the long-run; Hohwy 2017b).

If this account is on the right track, the generative model can be construed as a kind of (Bayesian) filter (Friston et al. 2010b) that strips sensory signals of their higher-frequency components as they are passed up the hierarchy. In conjunction with the 'horizontal' temporal processing described above (which can likewise be understood in terms of noncausal filtering or smoothing, where past and future state estimates are updated in light of novel sensory data; Friston et al. 2017a), this 'vertical' filtering scheme enables the organism to form reliable higher-order representations of the slowly-evolving statistical regularities underlying rapid sensory fluctuations. The organism is thus able to model the slow oscillatory dynamics embedded within the distal structure of its eco-niche (e.g., the diurnal temperature cycle), even though the particular sensory states through which these dynamics are accessed may vary considerably over time (e.g., temperature variation may be modulated by multiple interacting factors subtending multiple timescales—momentary occlusion of the sun, daily and seasonal weather cycles, climate change, etc.).

These dual facets of representational detachment help to explain not only how the hierarchical agent learns about invariant properties of an ever-changing environment, but also how it can exploit such regularities to its advantage. Circadian rhythms offer a particularly good example of how abstract representations of oscillatory dynamics foster adaptive behaviour in the context of environmental uncertainty.³⁷ Given a reliable model of how certain environmental properties are likely to evolve, the agent can form allostatic predictions that enable it to act in preparation for impending conditions, even if such expectations run contrary to current sensory evidence.

An interesting corollary of this view is the role of allostatic representations (e.g., circadian templates or programmes of activity) in compelling the agent to act 'as if' particular states of affairs obtain. Under certain conditions, such allostatic predictions amount to a kind of *false inference* about the hidden states that are currently in play. Although such predictions might be expected to engender actions that accumulate prediction error, the agent persists with them on account of their prior

³⁷ For discussion on the representational status of circadian rhythms, see Bechtel (2011) and Morgan (2018a, b).

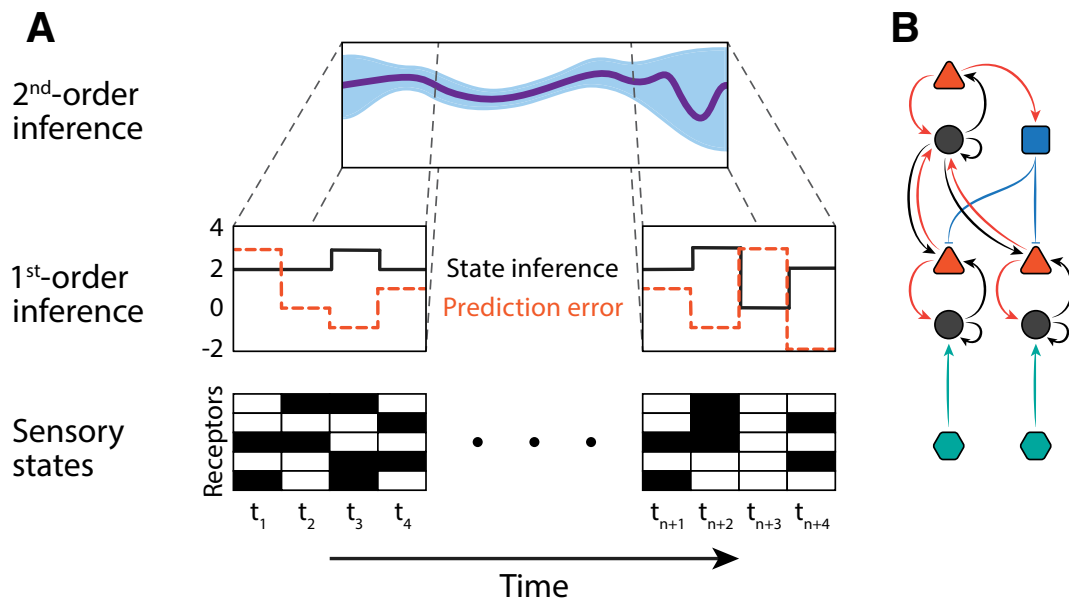


Fig. 2 Perceptual dynamics under hierarchical active inference. **a** In this illustration, the minimal active inference scheme has been augmented with a second-order perceptual inference level that tracks changes in the nutrient gradient over time. The purple function in the top panel indicates the agent's time-evolving estimate of ambient nutrient levels, which is derived from first-order sensory inferences (middle panels) on successive chemosensory receptor states (bottom panel raster plots; black cells indicate occupied receptor sites at time t). This function oscillates slowly as detected nutrient levels remain more or less stable over time, each incoming 'packet' of sensory information smoothly integrated within the broader temporal horizon of predicted and postdicted sensory states. The function begins to oscillate more rapidly when the organism experiences marked deviations from its expected states (right panels). This sudden volley of prediction error precipitates an increase in the precision on first-order prediction errors, enhancing the agent's perceptual sensitivity to environmental fluctuations. Increasing variability of sensory input also induces greater uncertainty about the trajectory of sensory states (as reflected in the broadening blue shading). **b** Schematic of a possible implementation of the hierarchical active inference scheme depicted in A. Sensory input from chemoreceptors (green hexagons) is received at the first processing level and compared to sensory expectations (black circles). Discrepancies between expected and actual input generate prediction errors (red triangles), which are passed up the hierarchy to the second processing level. Crucially, these prediction errors are modulated by precision estimates (blue square), which determine the 'gain' or influence ascribed to error signals (where high gain compels expectation units to conform with prevailing sensory evidence). Expected precision over first-order prediction errors is modulated in turn by second-order prediction error, which increases the gain on first-order errors. See Kanai et al. (2015), Parr and Friston (2018a), and Shipp (2016) for more detailed discussion of how such hierarchical schemes might be implemented in the brain. Figure reproduced from Corcoran et al. 2019 (CC BY 4.0)

precision, which causes conflicting sensory evidence to be downweighted or attenuated (Brown et al. 2013; Wiese 2017).

Returning to our earlier example, let us imagine that the hierarchical agent leverages its internal representation of diurnal temperature variation to schedule its activities to coincide with favourable environmental conditions. For instance, the organism might preemptively downregulate metabolic activity in preparation for nocturnal quiescence, irrespective of whether the ambient temperature has declined to an extent that would impair its metabolic efficiency. Likewise, the agent might begin to upregulate its activity around its usual time of 'awakening', despite the fact

that this routine provokes an elevated rate of energy expenditure on an usually chilly morning.

On first blush, this arrangement might seem suboptimal—surely the agent would be better off tuning its behaviour to *actual* environmental conditions, rather than relying on error-prone predictions? However, this would simply return us to the kind of closed-loop architecture of the minimal active inference agent; a creature incapable of distinguishing a genuine change in distal conditions from a transient deviation in its sensory states. In this sense our agent’s circadian gambit constitutes a more intelligent mode of regulation—armed with implicit knowledge of how state trajectories tend to evolve, the organism acts on the assumption that the future will roughly approximate the past, and treats transient deviations from this prescribed pattern as mere noise (i.e. the inherent uncertainty associated with stochastic processes).

Hence, although circadian rhythms might not guarantee ideal behaviour on shorter timescales, their adaptive value inheres in their ability to approximate the trajectory of homeostatically-relevant states over time. Such allostatic representations provide useful heuristics for guiding action—behaving in accordance with circadian predictions keeps the agent within the vicinity of its attracting set, thus affording a highly efficient means of reducing average uncertainty. Representations of this sort are insensitive to short-term fluctuations precisely because such transient dynamics (e.g., an unseasonably cold morning) are unlikely to afford information that improves its capacity to accurately predict future states. Circadian rhythms are therefore ‘robust’ to outlying or stochastic fluctuations in sensory data, thus constituting a reliable model of the underlying generative process.³⁸

In contrast to the minimal active inference agent, the hierarchical organism can exploit regularities in its environment to predict when and where it will be best placed to act, rather than responding reflexively to online sensory updates. Yet, while deep hierarchical architectures afford substantial advantages over the minimal scheme of Model 1, their capacity to reduce uncertainty through parameter estimation is most effective in a relatively stable world. Sudden alterations in environmental conditions (e.g., exchanging the European winter for the Australasian summer) require relatively long periods of reparameterisation, and may engender suboptimal, surprise-accruing behaviour in the interim. Flexible adaptation to novel (or rapidly-changing) situations requires generative models endowed with a temporal depth that transcends the hierarchical separation of fast and slow dynamics. We discuss such models next.

³⁸ The remarkable robustness of circadian oscillations is thrown into relief whenever one traverses several time-zones—a good example of how strongly-held (i.e. high-precision or ‘stubborn’; see Yon et al. 2019) allostatic expectations persist in the face of contradictory sensory evidence (i.e. the phase-shifted photoperiod and feeding schedule, to which the system eventually recalibrates; Asher and Sassone-Corsi 2015; Menaker et al. 2013).

Model 3: Counterfactual active inference

Our final model describes a biological agent equipped with a *temporally deep* model, which furnishes the ability to explicitly predict and evaluate the consequences of its policies. While this kind of generative model is undoubtedly the most complex and sophisticated of our three active inference schemes, it is also the most powerful, insofar as it allows the agent to perform *counterfactual* active inference.³⁹

Counterfactual active inference adds to the hierarchical processing of progressively deeper models through subjunctive processing: The agent can evaluate the expected free energy of alternative policies *under a variety of different contexts* before alighting on the best course of action (Friston 2018; Limanowski and Friston 2018). Our understanding of subjunctive processing draws on the Stalnaker-Lewis analysis of counterfactual conditionals, where the truth-conditions of a consequent are determined in relation to the *possible world* invoked by its antecedent (Lewis 1973b; Stalnaker 1968, see also Nute 1975; Sprigge 1970; Todd 1964).⁴⁰ In the context of active inference, counterfactual processing translates to the simulation of those sensory states that the organism *would* observe if it *were* to enact a certain policy *under a particular set of model parameters* (i.e. a possible world).

Our formulation of counterfactual inference implies two complementary processes, which we briefly introduce here. The first of these involves counterfactual inference on policies under spatiotemporally distal conditions. For example, the agent could reflect on a previous decision that precipitated a negative outcome, and consider how events might have unfolded differently (for better or worse) had it selected an alternative course of action (i.e. ‘retrospective’ inference). Similarly, the agent could envisage a scenario that it might encounter in the future, and imagine how various policies might play out under these circumstances (i.e. ‘prospective’ inference). This kind of counterfactual processing is useful for resolving uncertainty over the outcomes expected under various policies, and is integral to many sophisticated forms of cognitive processing (e.g., causal induction, mental time travel, mindreading, etc.; Buckner and Carroll 2007; Pezzulo et al. 2017; Schacter and Addis 2007; Suddendorf and Corballis 1997, 2007).

The second kind of uncertainty reduction mediated by counterfactual processing pertains to the arbitration of policies when the state of the world is itself ambiguous. This situation may arise due to uncertainty about the context that currently obtains (or relatedly, uncertainty over the consequences of policies within a particular context), or because the inhabited niche is inherently volatile (i.e. prone to fluctuate in ways that are relevant for the organism’s wellbeing, yet difficult to anticipate). Under such circumstances, counterfactual hypotheses may prove useful in two ways: (1) they may enable the agent to infer the policy that minimises (average) uncertainty

³⁹ For further discussion of counterfactual representation under predictive processing, see Clark (2016, Ch. 3), Friston et al. (2012b), Friston (2018), Palmer et al. (2015), Pezzulo et al. (2015) and Seth (2014, 2015).

⁴⁰ Note that our use of counterfactual semantics here is not intended to imply that cognition bears any necessary resemblance to linguistic processing; it is simply adopted as a convenient way of characterising the logic of model selection under active inference.

across a variety of possible worlds; (2) they may point towards ‘epistemic’ actions that help to disambiguate the *actual* state of the world (i.e. disclose the likelihood mapping that currently obtains), thus improving precision over policies.

As a brief illustration of counterfactual inference, let us consider an iteration of our *E. coli*-like creature that can evaluate the outcomes of its policies across several possible worlds. An organism sensitive to incident light could for instance run a counterfactual simulation for a possible world in which there is much scattered sunlight, and compare this to an alternative world featuring relatively little sunlight. If sunlight poses a threat to the bacterium (perhaps sun exposure causes the nutrient patch to dry up), tumbling constitutes a riskier strategy in the sun-dappled world. If it can order these possible worlds on the basis of their similarity to the actual world, then these counterfactual simulations could prove informative about the best action to take in a particular situation.⁴¹ Should the sun-dappled world turn out more similar to the actual world, then the organism would do well to confine its foraging activity to shady regions of the environment. The agent might consequently adapt its policies such that it tolerates gradient descent in the context of low incident light, only risking the *Tumble* policy when the nutrient supply is critically depleted.

Counterfactual processing enriches the generative model greatly, relative to the hierarchical organisation described in the previous section. Now there is wholly detached generative modelling of fine-grained elements of the prediction error landscape through simulated action; there is (Bayesian) model selection in terms of the best policy (i.e. minimising the free energy between the nutrient gradient simulated under a policy and the organism’s expected nutrient gradient; cf. Fitzgerald et al. 2014; Friston et al. 2016, 2017b; Parr and Friston 2018b); and there is processing that orders possible worlds (i.e. hypotheses entailed under competing model parameterisations) according to their comparative similarity to the actual world (where similarity may be cashed out in terms of representations of law-like relations (e.g., between nutrient gradient and sunlight) and particular matters of fact (e.g., amount of nutrient and sunlight); cf. Lewis 1973a, b, 1979). This contrasts sharply with the hierarchical agent, whose representational states are never completely detached from the content of its sensory flows, and whose active states are modulated gradually in response to reliable patterns of covariation.

More formally, counterfactual active inference rests on the ability to calculate the expected free energy of one’s policies. This is important for our analysis because the expected free energy of a policy can be decomposed into two terms—expected complexity and expected accuracy—which can be regarded as two kinds of uncertainty: *risk* and *ambiguity* (Friston et al. 2017a, b, d).⁴² Technically, risk constitutes

⁴¹ Interestingly, recent psychological evidence suggests that counterfactual scenarios deemed more similar to previously experienced events are perceived as more plausible and easier to envisage (i.e. simulate) than more distant alternatives (Stanley et al. 2017). This observation lends weight to the idea that humans evaluate competing counterfactual predictions in accordance with their proximity to actual states of affairs, where proximity or similarity might be cashed out in terms of (Bayesian) model evidence (see Fitzgerald et al. 2014).

⁴² Risk and ambiguity are also known as irreducible uncertainty and (parameter) estimation uncertainty, respectively (de Berker et al. 2016; Payzan-LeNestour and Bossaerts 2011). Note that uncertainty can be

a relative uncertainty (i.e. entropy) about predicted outcomes, relative to preferred outcomes, whereas ambiguity is a conditional uncertainty (i.e. entropy) about outcomes given their causes. More intuitively, risk can be understood as the probability of gaining some reward (e.g., finding a cookie) as a consequence of some action (e.g., reaching into a cookie jar), while ambiguity pertains to the fact that an observation might have come about in a variety of different ways (e.g., the cookie in my hand might have been given to me, stolen from the jar, etc.).⁴³ Counterfactual active inference agents need to consider both of these sources of uncertainty during policy selection. This is because resolving ambiguity will increase the agent's confidence about the process(es) responsible for generating observations, enabling it to calculate the risk (i.e. expected cost) associated with alternative courses of action.

With counterfactual inference at its disposal, the organism is potentially even better equipped to meet the demands of a complex and capricious environment.⁴⁴ Rather than engaging 'hard-wired' responses to current states (cf. Model 1), or 'soft-wired' responses to anticipated states (cf. Model 2), it can exploit *offline* computation of the likely consequences of different policies under various hypothetical conditions (Gärdenfors 1995; Grush 2004; Pezzulo 2008). This affords the opportunity to generate and test a wide variety of policies in the safety of its imagination, where actions that turn out to be too risky (or downright stupid) can be safely trialed and (hopefully) rejected (cf. Craik 1943, p. 61; Dennett 1995, pp. 375–376; Godfrey-Smith 1996, pp. 105–106). This capacity (or *competence*, see Williams 2018) to disengage from the present and undertake such 'thought experiments' confers a powerful mechanism for innovation, problem-solving, and (vicarious) learning—major

Footnote 42 (continued)

decomposed in various other ways, depending on the domain of interest (see for e.g., Bland and Schaefer 2012; Bradley and Drechsler 2014; Kozyreva and Hertwig 2019).

⁴³ This characterisation of risk and ambiguity is broadly consistent with descriptions in economics (e.g., Camerer and Weber 1992; Ellsberg 1961; Kahneman and Tversky 1979; Knight 1921) and neuroscience (e.g., Daw et al. 2005; Hsu et al. 2005; Huettel et al. 2006; Levy et al. 2010; Payzan-LeNestour and Bossaerts 2011; Preusschoff et al. 2008; for a review, see Bach and Dolan 2012). Importantly, these two sorts of uncertainty rest upon the precision (inverse variability) of the likelihood mapping between outcomes and hidden states—and transitions amongst hidden states that may or may not be under the creature's control. Technically, the first sort of precision relates to observation noise, while the second relates to system or state noise, i.e. volatility. Formally, volatility can be construed as the (inverse) precision over transition probabilities (i.e. confidence about the way hidden states evolve over time; Parr and Friston 2017; Parr et al. 2019; Sales et al. 2019; Vincent et al. 2019). This formulation suggests that volatile environments will tend to generate more surprising outcomes than stable environments, insofar as their states are apt to change in ways that are difficult to anticipate. Note that the term volatility is used differently in various contexts (see for e.g., Behrens et al. 2007; Bland and Schaefer 2012; Mathys et al. 2014).

⁴⁴ One caveat to this claim is that the (neuro)physiological mechanisms and cognitive operations required to enrich and exploit counterfactual predictive models may themselves engender additional costs (e.g., planning a new course of action requires time, energy, and effort; see Zénon et al. 2018). We assume that the costs incurred by such processes 'pay for themselves' over the long-run (or at least tend to on average), insofar as they enable the agent to exploit prior experience in ways that are conducive to adaptive behaviour (see Buzsáki et al. 2014; Pezzulo 2014; Pezzulo et al. 2017; Suddendorf et al. 2018). It is also worth pointing out that some of the costs engendered by counterfactual inference-supporting architectures may be mitigated by a variety of adaptive strategies (e.g., model updating during sleep, habitisation of behaviour under stable and predictable conditions; see Fitzgerald et al. 2014; Friston et al. 2017b; Hobson and Friston 2012; Pezzulo et al. 2016).

advantages in complex environments (Buzsáki et al. 2014; Mugan and MacIver 2019; Redish 2016).

The counterfactual active inference scheme described here implies additional degrees of organismic complexity that can be exploited to mitigate the impact of environmental uncertainty. The counterfactual agent is not only capable of ‘expecting the unexpected’ (inasmuch as it can countenance states of affairs that are unlikely under its current model of reality), but can prepare for it too—exploiting counterfactual hypotheses to formulate strategies for solving novel problems that might arise in the future (e.g., deciding what one should do in the event of sustaining a puncture while cycling to work). Moreover, the agent may organise its policy sets in ways that are sensitive to outcome contingencies, such that it can choose a backup policy if its initial plan is thwarted (e.g., being prepared to order the apple pie if the tiramisu has sold out). This ability to deftly switch between a subset of low-risk policies may confer a huge advantage under changing (or volatile) environmental conditions, where the time and effort required to re-evaluate a large array of policies from scratch could prove extremely costly.

Counterfactual processing is also valuable when the system is confronted with a sudden or sustained volley of prediction error. The counterfactual agent is able to interpret such signals as evidence that the hidden dynamics underwriting its sensory flows may have changed in some significant way (e.g., finding oneself confronted by oncoming traffic), and can draw on alternative possible models to evaluate which parameterisation affords the best explanation for the data at hand (cf. *parameter exploration*; Schwartenbeck et al. 2019). If the contingent relations structuring relevant environmental properties have indeed altered (e.g., realising one is visiting a country where people drive on the opposite side of the road), the agent will need to update its model so as to capture these novel conditions (see Sales et al. 2019). Failure to do so runs the risk of accruing further prediction error, since persisting with policies predicated on inaccurate (i.e. ‘out-of-date’) likelihood mappings may yield highly surprising outcomes.

One way to assess whether conditions or contexts have indeed changed is to engage in *epistemic action*, the final feature of counterfactual active inference we address here. Epistemic actions are active states that are sampled in order to acquire information about environmental contingencies (Friston et al. 2015b, 2016, 2017a, d).⁴⁵ When faced with the problem of identifying which model best captures the causal structure of the world, the agent can run simulations to infer the sensory flows each model predicts under a certain policy. The agent can then put these hypotheses to the test by sampling actions designed to arbitrate amongst competing predictions (Seth 2015). If the agent selects actions that are high in *epistemic value*, it

⁴⁵ For the purposes of this brief discussion, we limit the scope of epistemic action to instances where the organism actively intervenes on its environment in order to resolve uncertainty. It is worth noting, however, that the concept can also refer to *mental actions* or cognitive operations that reduce uncertainty (see for e.g., Metzinger 2017; Pezzulo et al. 2016; Pezzulo 2017). On this broader understanding, one might construe the different varieties of counterfactual processing described above as covert modes of epistemic action.

will observe outcomes that afford decisive evidence in favour of the model that best captures the current environmental regime.

The possibility of resolving ambiguity over the parameterisation of state–outcome contingencies through counterfactually-guided epistemic action also extends to ambiguity over policies. Here, the agent may run counterfactual simulations to infer actions that are likely to harvest information that clarifies the best policy to pursue.⁴⁶ These epistemic capabilities recapitulate the point that the policies of the counterfactual agent are not only scored with respect to risk-reduction or expected value (i.e. the extent to which they are expected to realise a *preferred* outcome), but also with respect to ambiguity-reduction or epistemic value (i.e. the extent to which they are expected to produce an *informative* outcome). Such epistemic actions are unavailable to the (merely) hierarchical agent, who can only reduce uncertainty over model parameters by slowly tuning its estimates to capture stable, enduring patterns of variation.⁴⁷

Two options for cognition

We began this paper with the lofty ambition of learning something about the nature and function of cognition, but have for the most part been careful to eschew talk of the cognitive or the mental. In this final section, we sketch out some of the broader implications of our analysis for the concept of biological cognition, and how the latter might be delimited from more general notions of life and adaptive plasticity.

As a precursory step, let us begin by considering how the three schematic models described in “[Biological regulation in an uncertain world](#)” might relate to real biological agents. One obvious strategy would be to map these architectures onto different taxonomic classes. For instance, one might construe the difference between these models as approximating the difference between relatively primitive organisms (like *E. coli* and other unicellular organisms), creatures with some degree of hierarchical depth (like reptiles or fish), and animals that demonstrate evidence of counterfactual sensitivity (like rodents; e.g., Redish 2016; Steiner and Redish 2014; Sweis et al. 2018; corvids; e.g., Bugnyar et al. 2016; Kabadayi and Osvath 2017; Raby et al. 2007; and primates; e.g., Abe and Lee 2011; Krupenye et al. 2016; Lee et al. 2005).

⁴⁶ Such activity is sometimes referred to as *epistemic foraging*, where the agent seeks out information about the way state transitions are likely to unfold (Friston et al. 2017d; Mirza et al. 2016; Parr and Friston 2017). For a nice example of epistemic foraging in wild dolphins, see Arranz et al. (2018).

⁴⁷ It is interesting to remark how epistemic action contributes to the practical utility of cognition as understood under the environmental complexity thesis. Following Dewey (1929), Godfrey-Smith (1996, pp. 116–120) notes that cognition is most likely to be useful in environments that comprise a mixture of regularity and unpredictability. Specifically, distal states should vary in ways that are a priori unpredictable (but worth knowing about), while maintaining a stable relationship with proximal states (see also Dunlap and Stephens 2016). The capacity to engage in epistemic action enhances the potential utility of cognition precisely insofar as it helps the agent to reduce uncertainty over this mapping, thus affording more precise knowledge (or novel insight; Friston et al. 2017b) about the state of the world and its possible alternatives.

This approach is immediately undermined however by the remarkable complexity evinced by (at least some) unicellular organisms. Bacteria like *E. coli* integrate information over a variety of sensory channels, modulate their metabolic and chemotactic activity in response to reliable environmental contingencies, and alternate policy preferences in a context-sensitive fashion (Ben-Jacob 2009; Freddolino and Tavazoie 2012; Hennessey et al. 1979; Mitchell et al. 2009; Salman and Libchaber 2007; Tagkopoulos et al. 2008; Tang and Marshall 2018; see also van de Cruys 2017, for discussion from a predictive processing perspective). Although this does not rule out the possible existence of minimal active inference agents, it might suggest that *all* extant lifeforms instantiate some form of allostatic architecture. This raises the question of whether meaningful distinctions can be drawn in terms of hierarchical organisation (e.g. shallow vs. deep hierarchies), and whether such distinctions can be systematically mapped to particular functional profiles (e.g., capacities for learning and adaptive flexibility).

It might also be tempting to think of our model organisms as exemplifying creatures that are more or less evolved or adapted to their environment. Undoubtedly, the counterfactual agent comprises a more complex information-processing architecture than its minimal active inference counterpart, one equipped with a much greater capacity for flexible, selective adaptation to the vicissitudes wrought by uncertainty. However, we must be careful not to conflate adaptation to a specific set of environmental properties with adaptation to environmental complexity per se. On both the environmental complexity thesis and the free energy principle, organisms are adapted to their environments to the extent that they successfully track and neutralise *ecologically-relevant* sources of uncertainty (cf. ‘frugal’ generative models; Battieri and Buckley 2017; Clark 2015). This means that organisms comprising radically divergent degrees of functional complexity can in principle constitute equally good models of the same environment, assuming they are equally capable of acting in ways that minimise the conditional entropy over their sensory states.

Finally, given that the free energy principle conceives of all biological agents as being engaged in the same essential activity (i.e. the singular project of minimising free energy, maximising self-evidence, and thus conserving self-organisation over time), one might question whether there really are any substantive differences to be found between the levels of our three-tiered scheme. In conjunction with the argument presented in the previous paragraph, it might seem that these architectures differ from one another in a fairly superficial way: They simply illustrate alternative solutions to the fundamental problem of uncertainty reduction over time.

This point notwithstanding, we believe that the distinct functional capacities we have ascribed to these models carry important implications about the origins and limits of cognition. The fact that all three architectures are afforded equal footing by the free energy principle does not speak against this view—despite its neuroscientific origins (Friston 2002, 2003, 2005), the free energy principle makes no explanatory commitments to cognition per se; it simply imposes certain formal constraints on the sort of functional organisation a cognitive system must realise in order to resist entropy. This marks a significant distinction from the environmental complexity thesis, which on Godfrey-Smith’s telling limits its explanatory scope to the *subset* of living organisms that count as cognitive agents.

Put differently, the free energy principle is *neutral* on the ontological relation between life and cognition (*pace* Kirchhoff and Froese 2017). The environmental complexity thesis, on the other hand, endorses a *weak continuity* (“Anything that has a mind is alive, although not everything that is alive has a mind”; Godfrey-Smith 1996, p. 72) without specifying a principled way of demarcating the boundary between the cognitive and the non-cognitive.⁴⁸ We suggest this boundary can be located at the nexus between hierarchical and counterfactual forms of active inference. This would mean that only those biological systems capable of engaging in fully detached modes of representation, and of exploiting such representations for the purposes of uncertainty reduction, count as cognitive agents.⁴⁹

Associating cognition with counterfactual active inference might strike some as unduly restrictive, limiting category membership to humans and only the most intelligent of mammals and birds (for instance). It is important to bear in mind, however, that our construal of counterfactual processing is a formal one; many kinds of animals are likely to exploit counterfactual inferences in ways that enable them to learn about the world and make sensible (uncertainty-reducing) decisions. Some of these processing architectures might turn out to be highly impoverished compared to the rich counterfactual capacities at our own disposal (cf. Carruthers 2004), but we consider this difference a matter of degree, not kind.

Notably, our counterfactual criterion does not exclude such organisms as bacteria, protists, and plants from the cognitive domain by *fiat*. If clever empirical studies were to reveal that *E. coli* (for example) proactively solicit ambiguity-reducing information to plan their future chemotactic forays, this would afford compelling evidence they constitute cognitive agents. However, as pointed out in recent debates about future-oriented cognition in non-human animals, seemingly complex patterns of behaviour do not always licence the attribution of complex representational or inferential capacities (Redshaw and Bulley 2018; Suddendorf and Redshaw 2017; see Mikhalevich et al. 2017, for an environmental complexity-inflected counterargument). If empirical observations can be parsimoniously explained by appeal to such allostatic mechanisms as information integration (Read et al. 2015) and elemental

⁴⁸ Godfrey-Smith thus rejects *strong continuity*, the view that “[l]ife and mind have a common abstract pattern or set of basic organizational properties. [...] Mind is literally life-like” (1995, p. 320, emphasis in original). Evan Thompson (2007) has defended a position similar to this (‘deep continuity’), albeit with the addition of an existential-phenomenological supplement (for discussion, see Wheeler 2011). This view inherits from Maturana’s canonical account of autopoiesis, where one finds the strongest expression of life–mind continuity: “Living systems *are* cognitive systems, and living as a process *is* a process of cognition” (Maturana and Varela 1980, p. 13, emphasis added; see also Heschl 1990).

⁴⁹ It is perhaps worth noting that other scholars have used the criterion of “detachment” (or “decouplability”) to distinguish representational versus non-representational agents, rather than cognitive versus non-cognitive agents (cf. Clark and Grush 1999; Grush 2004). Without digressing into a discussion of the relationship between representational and cognitive systems, we remark that our view conceives of cognition as a computational architecture that engages in a particular subset of representational operations—i.e. the generation, manipulation, and evaluation of counterfactual model predictions. These operations are situated within a broader class of uncertainty-resolving processes, including the homeostatic and allostatic representational schemes outlined in “Biological regulation in an uncertain world”.

learning (Giurfa 2013; Perry et al. 2013), admittance to the cognitive domain ought to be withheld.

An alternative (and increasingly popular) approach would be to ascribe some form of ‘minimal’ or ‘proto-cognitive’ status to bacteria, plants, and other aneural organisms (Ben-Jacob 2009; Calvo Garzón and Keijzer 2011; Gagliano 2015; Godfrey-Smith 2016a, b; Lyon 2015, 2019; Segundo-Ortin and Calvo 2019; Smith-Ferguson and Beekman 2019; van Duijn et al. 2006; for a dissenting view, see Adams 2018). Such terms might seem appealing in light of the mounting body of research claiming that many ‘simple’ organisms engage in primitive or precursory forms of cognitive activity (Baluška and Levin 2016; Levin et al. 2017; Tang and Marshall 2018). Granting such cases do indeed demonstrate genuine instances of learning, memory, decision-making, and so on, it seems only the staunchest of neuro-chauvinists would persist in denying the cognitive status of such organisms.

While we cannot do justice to this complex topic here, a few remarks are in order. First, we should acknowledge that there may be few substantive differences between the kinds of organisms we designate as hierarchical or allostatic agents, and the biological systems Godfrey-Smith and others would identify as exhibiting ‘minimal’ or ‘proto-cognitive’ capacities (e.g., Godfrey-Smith 2002, 2016b).⁵⁰ Both categories imply systems that track relevant states in their (internal and external) environments, and exploit this information to adaptively regulate their activity. Both categories also imply some form of evolutionary precedence over ‘fully-fledged’ cognitive agents—cognition ‘proper’ builds on the foundations laid by allostatic/proto-cognitive architectures.

The problem with such terminology is that it implies the ascription of some form of cognitive capacity, while remaining opaque as to its precise relation to ‘full-blown’ cognition—including the reason for its segregation from the latter (see Lyon 2019, for an extended critique). Is there some fundamental cognitive ingredient that proto-cognition lacks, or is it simply a scaled-down, severely degraded version of (say) animal cognition? If the latter, is the distinction between proto- and ‘genuine’ cognition marked by a critical boundary, or is the difference gradual and indeterminate? Godfrey-Smith explicitly endorses some variety of the latter view, frequently remarking that cognition ‘shades-off’ into other biological processes. But if proto-cognitive organisms ultimately fail to qualify as cognitive agents,⁵¹ such talk may obscure a fundamental *discontinuity*.

We take it that the capacity for counterfactual processing marks the subtle but significant functional boundary hinted at in Godfrey-Smith’s analysis. This proposal is—in most cases—stricter than other criteria often mentioned in the debate about minimal cognition: it implies that organisms that only engage in allostatic regulation (sometimes requiring forms of learning, memory, and decision-making) would not

⁵⁰ ‘Minimal cognition’ is perhaps more closely associated with a rather different set of philosophical views than those espoused by Godfrey-Smith (e.g., anti-representationalism, situated and embodied cognition; Barandiaran and Moreno 2006; Beer 2003; van Duijn et al. 2006). We take the main thrust of our argument to be equally applicable to these positions.

⁵¹ When pressed, Godfrey-Smith seems to hold this view: “I do *not* claim that bacteria exhibit cognition; this is *at most* a case of proto-cognition” (2002, p. 223, emphasis added).

necessarily qualify as cognitive agents. Of course, testing which organisms meet this counterfactual criterion remains an important conceptual and empirical challenge.

In this respect, our proposed definition is not neuro-chauvinistic, but is focussed rather on a functional (computationally-grounded) definition of cognition that can be met—at least in principle—by many different kinds of organisms. On this view, a minimally cognitive agent is a minimally counterfactual agent—an organism that not only learns about itself and its environment, but imagines them anew. If we are wrong, and sophisticated forms of cognitive activity simply emerge as allostatic processing schemes become increasingly more powerful and hierarchically elaborate, then a single dimension along which cognition ‘shades off’ into primitive forms of sensorimotor control and metabolic regulation would seem the better option.

Acknowledgements AWC is supported by an Australian Government Research Training Program (RTP) scholarship. JH is supported by the Australian Research Council (DP160102770, DP190101805). This research has received funding from the European Union’s Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2 to GP). We would like to thank participants at the *Science of the Self* research forum and the *22nd Annual Meeting of the Association for the Scientific Study of Consciousness* for feedback on earlier presentations of this work. We also wish to thank Louise Kyriaki, Dan Williams, members of the Cognition & Philosophy Lab—especially Stephen Gadsby, Andy McMilliam, Kelsey Perrykkad, and Iwan Williams—and two anonymous reviewers for insightful comments on earlier versions of this manuscript.

References

- Abe H, Lee D (2011) Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron* 70(4):731–741
- Adams F (2018) Cognition wars. *Stud Hist Philos Sci* 68:20–30
- Adams RA, Shipp S, Friston KJ (2013) Predictions not commands: active inference in the motor system. *Brain Struct Funct* 218(3):611–643
- Ainley V, Apps MAJ, Fotopoulou A, Tsakiris M (2016) ‘Bodily precision’: a predictive coding account of individual differences in interoceptive accuracy. *Philos Trans R Soc B* 371(20160003):1–9
- Allen M, Friston KJ (2018) From cognitivism to autopoiesis: towards a computational framework for the embodied mind. *Synthese* 195(6):2459–2482
- Allen M, Tsakiris M (2018) The body as first prior: Interoceptive predictive processing and the primacy of self-models. In: Tsakiris M, De Preester H (eds) *The interoceptive mind: from homeostasis to awareness*. Oxford University Press, Oxford, pp 27–45
- Allen M, Levy A, Parr T, Friston KJ (2019) In the body’s eye: the computational anatomy of interoceptive inference. *bioRxiv*
- Andrews BW, Yi T-M, Iglesias PA (2006) Optimal noise filtering in the chemotactic response of *Escherichia coli*. *PLoS Comput Biol* 2(11):e154
- Arranz P, Benoit-Bird KJ, Southall BL, Calambokidis J, Friedlaender AS, Tyack PL (2018) Risso’s dolphins plan foraging dives. *J Exp Biol* 221(4):jeb165209
- Ashby WR (1940) Adaptiveness and equilibrium. *Br J Psychiatry* 86(362):478–483
- Ashby WR (1956) *An introduction to cybernetics*. Chapman & Hall Ltd, London
- Ashby WR (1958) Requisite variety and its implications for the control of complex systems. *Cybernetica* 1(2):83–99
- Ashby WR (1960) *Design for a brain: The origin of adaptive behaviour*, 2nd edn. Chapman & Hall Ltd., London
- Asher G, Sassone-Corsi P (2015) Time for food: the intimate interplay between nutrition, metabolism, and the circadian clock. *Cell* 161(1):84–92

- Attias H (2003) Planning by probabilistic inference. In: Bishop CM, Frey BJ (eds) *Proceedings of the ninth international conference on artificial intelligence and statistics*. Society for Artificial Intelligence and Statistics, New Jersey
- Bach DR, Dolan RJ (2012) Knowing how much you don't know: a neural organization of uncertainty estimates. *Nat Rev Neurosci* 13:572–586
- Badcock PB, Davey CG, Whittle S, Allen NB, Friston KJ (2017) The depressed brain: an evolutionary systems theory. *Trends Cognitive Sci* 21(3):182–194
- Badre D (2008) Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cognitive Sci* 12(5):193–200
- Bailey SM, Udoh US, Young ME (2014) Circadian regulation of metabolism. *J Endocrinol* 222(2):R75–R96
- Baltieri M, Buckley CL (2017) An active inference implementation of phototaxis. In: Knibbe C, Beslon G, Parsons D, Misevic JR-C, Bredèche N, Hassas S, Simonin O, Soula H (eds) *Proceedings of ECAL 2017: the 14th European conference on artificial life*. MIT Press, Cambridge, pp 36–43
- Baluška F, Levin M (2016) On having no head: cognition throughout biological systems. *Front Psychol* 7:902
- Barandiaran XE, Moreno A (2006) On what makes certain dynamical systems cognitive: a minimally cognitive organization program. *Adapt Behav* 14(2):171–185
- Barrett LF, Simmons WK (2015) Interoceptive predictions in the brain. *Nat Rev Neurosci* 16(7):419–429
- Barrett LF, Quigley KS, Hamilton P (2016) An active inference theory of allostasis and interoception in depression. *Philos Trans R Soc B* 371(20160011):1–17
- Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding. *Neuron* 76(4):695–711
- Bauman DE (2000) Regulation of nutrient partitioning during lactation: Homeostasis and homeorhesis revisited. In: Cronjé PB (ed) *Ruminant physiology: digestion, metabolism, growth and reproduction*, chapter 18. CABI Publishing, New York, pp 311–328
- Bechtel W (2011) Representing time of day in circadian clocks. In: Newen A, Bartels A, Jung E-M (eds) *Knowledge and representation*, Chapter 7. CSLI Publications, Stanford, pp 129–162
- Beer RD (2003) The dynamics of active categorical perception in an evolved model agent. *Adapt Behav* 11(4):209–243
- Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. *Nat Neurosci* 10(9):1214–1221
- Ben-Jacob E (2009) Learning from bacteria about natural information processing. *Ann N Y Acad Sci* 1178:78–90
- Berg HC, Brown DA (1972) Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. *Nature* 239(5374):500–504
- Bernard C (1974) *Lectures on the phenomena of life common to animals and plants*. American Lecture Series. Charles C. Thomas Pub Ltd, Springfield
- Berntson GG, Cacioppo JT (2000) From homeostasis to alldynamic regulation. In: Cacioppo JT, Tassinary LG, Berntson GG (eds) *Handbook of psychophysiology*, Chapter 17, 2nd edn. Cambridge University Press, Cambridge, pp 459–481
- Berridge KC (2004) Motivation concepts in behavioral neuroscience. *Physiol Behav* 81(2):179–209
- Bich L, Mossio M, Ruiz-Mirazo K, Moreno A (2016) Biological regulation: controlling the system from within. *Biol Philos* 31(2):237–265
- Birkhoff GD (1931) Proof of the ergodic theorem. *Proc Natl Acad Sci* 17(12):656–660
- Bland AR, Schaefer A (2012) Different varieties of uncertainty in human decision-making. *Front Neurosci* 6:85
- Blei DM, Kucukelbir A, McAuliffe JD (2017) Variational inference: a review for statisticians. *J Am Stat Assoc* 112(518):859–877
- Bogacz R (2017) A tutorial on the free-energy framework for modelling perception and learning. *J Math Psychol* 76:198–211
- Botvinick M, Toussaint M (2012) Planning as inference. *Trends Cogn Sci* 16(10):485–488
- Bradley R, Drechsler M (2014) Types of uncertainty. *Erkenntnis* 79(6):1225–1248
- Brown H, Adams RA, Parees I, Edwards M, Friston KJ (2013) Active inference, sensory attenuation and illusions. *Cogn Process* 14(4):411–427
- Bruineberg J, Rietveld E, Parr T, van Maanen L, Friston KJ (2018) Free-energy minimization in joint agent-environment systems: a niche construction perspective. *J Theor Biol* 455:161–178

- Buckley CL, Chang SK, McGregor S, Seth AK (2017) The free energy principle for action and perception: a mathematical review. *J Math Psychol* 81:55–79
- Buckner RL, Carroll DC (2007) Self-projection and the brain. *Trends Cogn Sci* 11(2):49–57
- Bugnyar T, Reber SA, Buckner C (2016) Ravens attribute visual access to unseen competitors. *Nat Commun* 7:10506
- Burdakov D (2019) Reactive and predictive homeostasis: roles of orexin/hypocretin neurons. *Neuropharmacology* 154:61–67
- Buzsáki G, Peyrache A, Kubie J (2014) Emergence of cognition from action. *Cold Spring Harb Symp Quant Biol* 79:41–50
- Cabanac M (1971) Physiological role of pleasure. *Science* 173(4002):1103–1107
- Cabanac M (2006) Adjustable set point: to honor Harold T. Hammel. *J Appl Physiol* 100(4):1338–1346
- Calvo P, Friston KJ (2017) Predicting green: really radical (plant) predictive processing. *J R Soc Interface* 14(20170096):1–11
- Calvo Garzón P, Keijzer F (2011) Plants: adaptive behavior, root-brains, and minimal cognition. *Adapt Behav* 19(3):155–171
- Camerer C, Weber M (1992) Recent developments in modeling preferences: uncertainty and ambiguity. *J Risk Uncertain* 5:325–370
- Campbell JO (2016) Universal Darwinism as a process of Bayesian inference. *Front Syst Neurosci* 10:49
- Cannon WB (1914) The emergency function of the adrenal medulla in pain and the major emotions. *Am J Physiol* 33(2):356–372
- Cannon WB (1915) Bodily changes in pain, hunger, fear and rage: an account of recent researches into the function of emotional excitement. D. Appleton and Company, New York
- Cannon WB (1929) Organization for physiological homeostasis. *Physiol Rev* 9(3):399–431
- Cannon WB (1939) The wisdom of the body: revised and, enlarged edn. W. W. Norton & Company Inc., New York
- Carruthers P (2004) On being simple minded. *Am Philos Q* 41(3):205–220
- Clark A (2015) Radical predictive processing. *South J Philos* 53:3–27
- Clark A (2016) Surfing uncertainty: prediction, action, and the embodied mind. Oxford University Press, Oxford
- Clark A (2017) How to knit your own markov blanket: resisting the second law with metamorphic minds. In: Metzinger T, Wiese W (eds) *Philosophy and predictive processing*, Chapter 3. MIND Group, Frankfurt am Main, pp 1–19
- Clark A (2018) A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenol Cogn Sci* 17(3):521–534
- Clark A, Grush R (1999) Toward a cognitive robotics. *Adapt Behav* 7(1):5–16
- Conant RC, Ashby WR (1970) Every good regulator of a system must be a model of that system. *Int J Syst Sci* 1(2):89–97
- Corcoran AW (2019) Cephalopod molluscs, causal models, and curious minds. *Anim Sentience* 4(26):13
- Corcoran AW, Hohwy J (2018) Allostasis, interoception, and the free energy principle: feeling our way forward. In: Tsakiris M, De Preester H (eds) *The interoceptive mind: from homeostasis to awareness*, Chapter 15. Oxford University Press, Oxford, pp 272–292
- Corcoran AW, Pezzulo G, Hohwy J (2018) Commentary: Respiration-entrained brain rhythms are global but often overlooked. *Front Syst Neurosci* 12:25
- Corcoran AW, Pezzulo G, Hohwy J (2019) From allostatic agents to counterfactual cognisers: active inference, biological regulation, and the origins of cognition. Preprints, 2019110083
- Craig AD (2009) How do you feel—now? The anterior insula and human awareness. *Nat Rev Neurosci* 10(1):59–70
- Craik K (1943) *The nature of explanation*. Cambridge University Press, Cambridge
- Crauel H, Flandoli F (1994) Attractors for random dynamical systems. *Probab Theory Relat Fields* 100:365–393
- Critchley HD, Harrison NA (2013) Visceral influences on brain and behavior. *Neuron* 77(4):624–638
- Dampney RAL (2016) Central neural control of the cardiovascular system: current perspectives. *Adv Physiol Educ* 40(3):283–296
- Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 8(12):1704–1711
- de Berker AO, Rutledge RB, Mathys C, Marshall L, Cross GF, Dolan RJ, Bestmann S (2016) Computations of uncertainty mediate acute stress responses in humans. *Nat Commun* 7:10996

- de Vries B, Friston KJ (2017) A factor graph description of deep temporal active inference. *Front Comput Neurosci* 11:95
- Degaute JP, van de Borne P, Linkowski P, Van Cauter E (1991) Quantitative analysis of the 24-hour blood pressure and heart rate patterns in young men. *Hypertension* 18(2):199–210
- Dennett DC (1987) *The intentional stance*. MIT Press, Cambridge
- Dennett DC (1995) *Darwin's dangerous idea: evolution and the meanings of life*. Penguin Books Ltd, London
- Dewey J (1929) *Experience and nature*. George Allen & Unwin Ltd, London
- Dolan RJ, Dayan P (2013) Goals and habits in the brain. *Neuron* 80(2):312–325
- Dunlap AS, Stephens DW (2016) Reliability, uncertainty, and costs in the evolution of animal learning. *Curr Opin Behav Sci* 12:73–79
- Dworkin BR (1993) *Learning and physiological regulation*. University of Chicago Press, Chicago
- Dyar KA, Lutter D, Artati A, Ceglia NJ, Liu Y, Armenta D, Jastroch M, Schneider S, de Mateo S, Cervantes M, Abbondante S, Tognini P, Orozco-Solis R, Kinouchi K, Wang C, Swerdlhoff R, Nadeef S, Masri S, Magistretti P, Orlando V, Borrelli E, Uhlenhaut NH, Baldi P, Adamski J, Tschöp MH, Eckel-Mahan K, Sassone-Corsi P (2018) Atlas of circadian metabolism reveals system-wide coordination and communication between clocks. *Cell* 174(6):1571–1585
- Elias P (1955) Predictive coding—part I. *IRE Trans Inf Theory* 1(1):16–24
- Ellsberg D (1961) Risk, ambiguity, and the Savage axioms. *Q J Econ* 75(4):643–669
- Evans DJ, Searles DJ (1994) Equilibrium microstates which generate second law violating steady states. *Phys Rev E* 50(2):1645–1648
- Evans DJ, Searles DJ (2002) The fluctuation theorem. *Adv Phys* 51(7):1529–1585
- Falke JJ, Bass RB, Butler SL, Chervitz SA, Danielson MA (1997) The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annu Rev Cell Dev Biol* 13:457–512
- Fernö A, Pitcher TJ, Melle W, Nøttestad L, Mackinson S, Hollingworth C, Misund OA (1998) The challenge of the herring in the Norwegian sea: making optimal collective spatial decisions. *Sarsia* 83(2):149–167
- Feynman RP (1972) *Statistical mechanics: a set of lectures*. W. A. Benjamin Inc, Reading
- FitzGerald THB, Dolan RJ, Friston KJ (2014) Model averaging, optimal inference, and habit formation. *Front Hum Neurosci* 8(457):1–11
- FitzGerald THB, Dolan RJ, Friston KJ (2015) Dopamine, reward learning, and active inference. *Front Comput Neurosci* 9(136):1–16
- Fotopoulou A, Tsakiris M (2017) Mentalizing homeostasis: the social origins of interoceptive inference. *Neuropsychanalysis* 19(1):3–28
- Freddolino PL, Tavazoie S (2012) Beyond homeostasis: a predictive-dynamic framework for understanding cellular behavior. *Annu Rev Cell Dev Biol* 28:363–384
- Friston KJ (2002) Functional integration and inference in the brain. *Prog Neurobiol* 68(2):113–143
- Friston KJ (2003) Learning and inference in the brain. *Neural Netw* 16(9):1325–1352
- Friston KJ (2005) A theory of cortical responses. *Philos Trans R Soc B* 360(1456):815–836
- Friston KJ (2008) Hierarchical models in the brain. *PLoS Comput Biol* 4(11):e1000211
- Friston KJ (2009) The free-energy principle: a rough guide to the brain? *Trends Cogn Sci* 13(7):293–301
- Friston KJ (2010) The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11(2):127–138
- Friston KJ (2011) Embodied inference: or “I think therefore I am, if I am what I think”. In: Tschacher W, Bergomi C (eds) *The implications of embodiment: cognition and communication*. Imprint Academic, Exeter, pp 89–125
- Friston KJ (2012a) A free energy principle for biological systems. *Entropy* 14(11):2100–2121
- Friston KJ (2012b) Policies and priors. In: Gutkin B, Ahmed SH (eds) *Computational neuroscience of drug addiction*, Springer series in computational neuroscience 10. Springer, New York, pp 237–283
- Friston KJ (2013) Life as we know it. *J R Soc Interface* 10(86):20130475
- Friston KJ (2017) Self-evidencing babies: commentary on “Mentalizing homeostasis: the social origins of interoceptive inference” by Fotopoulou & Tsakiris. *Neuropsychanalysis* 19(1):43–47
- Friston KJ (2018) Am I self-conscious? (Or does self-organisation entail self-consciousness?). *Front Psychol* 9:579
- Friston KJ, Ao P (2012) Free energy, value, and attractors. *Comput Math Methods Med* 937860
- Friston KJ, Kiebel S (2009) Predictive coding under the free-energy principle. *Philos Trans R Soc B* 364(1521):1211–1221
- Friston KJ, Stephan KE (2007) Free-energy and the brain. *Synthese* 159(3):417–458

- Friston KJ, Kilner J, Harrison L (2006) A free energy principle for the brain. *J Physiol Paris* 100(1–3):70–87
- Friston KJ, Mattout J, Trujillo-Barreto N, Ashburner J, Penny WD (2007) Variational free energy and the Laplace approximation. *NeuroImage* 34(1):220–234
- Friston KJ, Daunizeau J, Kiebel SJ (2009) Reinforcement learning or active inference? *PLoS ONE* 4(7):e6421
- Friston KJ, Daunizeau J, Kilner J, Kiebel SJ (2010a) Action and behavior: a free-energy formulation. *Biol Cybern* 102(3):227–260
- Friston KJ, Stephan KE, Li B, Daunizeau J (2010b) Generalised filtering. *Math Problems Eng* 3:621670
- Friston KJ, Adams RA, Montague R (2012a) What is value—accumulated reward or evidence? *Front Neurobot* 6:11
- Friston KJ, Adams RA, Perrinet L, Breakspear M (2012b) Perceptions as hypotheses: saccades as experiments. *Front Psychol* 3(151):1–20
- Friston KJ, Breakspear M, Deco G (2012c) Perception and self-organized instability. *Front Comput Neurosci* 6(44):1–19
- Friston KJ, Samothrakis S, Montague R (2012d) Active inference and agency: optimal control without cost functions. *Biol Cybern* 106(8–9):523–541
- Friston KJ, Thornton C, Clark A (2012e) Free-energy minimization and the dark-room problem. *Front Psychol* 3:130
- Friston KJ, Schwartenbeck P, FitzGerald T, Moutoussis M, Behrens T, Dolan RJ (2013) The anatomy of choice: active inference and agency. *Front Hum Neurosci* 7(598):1–18
- Friston KJ, Levin M, Sengupta B, Pezzulo G (2015a) Knowing one's place: a free-energy approach to pattern regulation. *J R Soc Interface* 12(20141383):1–12
- Friston KJ, Rigoli F, Ognibene D, Mathys CD, Fitzgerald T, Pezzulo G (2015b) Active inference and epistemic value. *Cogn Neurosci* 6(4):187–224
- Friston KJ, FitzGerald T, Rigoli F, Schwartenbeck P, O'Doherty J, Pezzulo G (2016) Active inference and learning. *Neurosci Biobehav Rev* 68:862–879
- Friston KJ, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G (2017a) Active inference: a process theory. *Neural Comput* 29(1):1–49
- Friston KJ, Lin M, Frith CD, Pezzulo G, Hobson JA, Ondobaka S (2017b) Active inference, curiosity and insight. *Neural Comput* 29(10):2633–2683
- Friston KJ, Parr T, de Vries B (2017c) The graphical brain: belief propagation and active inference. *Netw Neurosci* 1(4):381–414
- Friston KJ, Rosch R, Parr T, Price C, Bowman H (2017d) Deep temporal models and active inference. *Neurosci Biobehav Rev* 77:388–402
- Friston KJ, Parr T, Zeidman P (2018) Bayesian model reduction. [arXiv:1805.07092](https://arxiv.org/abs/1805.07092)
- Fuster JM (2001) The prefrontal cortex—an update: time is of the essence. *Neuron* 30(2):319–333
- Fuster JM (2004) Upper processing stages of the perception–action cycle. *Trends Cogn Sci* 8(4):143–145
- Gagliano M (2015) In a green frame of mind: perspectives on the behavioural ecology and cognitive nature of plants. *AoB Plants* 7:75
- Gärdenfors P (1995) Cued and detached representations in animal cognition. *Behav Proc* 35:263–273
- Ginty AT, Kraynak TE, Fisher JP, Gianaros PJ (2017) Cardiovascular and autonomic reactivity to psychological stress: neurophysiological substrates and links to cardiovascular disease. *Auton Neurosci Basic Clin* 207:2–9
- Giurfa M (2013) Cognition with few neurons: higher-order learning in insects. *Trends Neurosci* 36(5):285–294
- Godfrey-Smith P (1995) Spencer and Dewey on life and mind. In: Boden MA (ed) *The philosophy of artificial life*, Oxford Readings in Philosophy, chapter 12. Oxford University Press, Oxford, pp 314–331
- Godfrey-Smith P (1996) Complexity and the function of mind in nature. *Cambridge Studies in Philosophy and Biology*. Cambridge University Press, Cambridge
- Godfrey-Smith P (2002) Environmental complexity and the evolution of cognition. In: Sternberg RJ, Kaufman JC (eds) *The evolution of intelligence*, Chapter 10. Lawrence Erlbaum Associates Inc, Mahwah, pp 223–250
- Godfrey-Smith P (2016a) Individuality, subjectivity, and minimal cognition. *Biol Philos* 31(6):775–796
- Godfrey-Smith P (2016b) Mind, matter, and metabolism. *J Philos* 113(10):481–506
- Goodwin GM, McCloskey DI, Mitchell JH (1972) Cardiovascular and respiratory responses to changes in central command during isometric exercise at constant muscle tension. *J Physiol* 226(1):173–190

- Grush R (2004) The emulation theory of representation: motor control, imagery, and perception. *Behav Brain Sci* 27(3):377–442
- Gu X, Hof PR, Friston KJ, Fan J (2013) Anterior insular cortex and emotional awareness. *J Comp Neurol* 521(15):3371–3388
- Hennessey TM, Rucker WB, McDiarmid CG (1979) Classical conditioning in paramecia. *Anim Learn Behav* 7(4):417–423
- Heschl A (1990) $L=C$: a simple equation with astonishing consequences. *J Theor Biol* 145:13–40
- Hobson JA, Friston KJ (2012) Waking and dreaming consciousness: neurobiological and functional considerations. *Prog Neurobiol* 98(1):82–98
- Hohwy J (2013) *The predictive mind*. Oxford University Press, Oxford
- Hohwy J (2016) The self-evidencing brain. *Noûs* 50(2):259–285
- Hohwy J (2017a) How to entrain your evil demon. In: Metzinger T, Wiese W (eds) *Philosophy and predictive processing*, Chapter 2. MIND Group, Frankfurt am Main, pp 1–15
- Hohwy J (2017b) Priors in perception: top-down modulation, Bayesian perceptual learning rate, and prediction error minimization. *Conscious Cogn* 47:75–85
- Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF (2005) Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310(5754):1680–1683
- Huang Y, Rao RPN (2011) Predictive coding. *Wiley Interdiscip Rev Cogn Sci* 2(5):580–593
- Huettel SA, Stowe CJ, Gordon EM, Warner BT, Platt ML (2006) Neural signatures of economic preferences for risk and ambiguity. *Neuron* 49(5):765–775
- Iodice P, Porciello G, Bufalari I, Barca L, Pezzulo G (2019) An interoceptive illusion of effort induced by false heart-rate feedback. *Proc Natl Acad Sci* 116(28):13897–13902
- Kabadayi C, Osvath M (2017) Ravens parallel great apes in flexible planning for tool-use and bartering. *Science* 357(6347):202–204
- Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47(2):263–292
- Kalman RE (1960) A new approach to linear filtering and prediction problems. *J Basic Eng* 82(1):35–45
- Kalman RE, Bucy RS (1961) New results in linear filtering and prediction theory. *J Basic Eng* 83(1):95–108
- Kanai R, Komura Y, Shipp S, Friston KJ (2015) Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philos Trans R Soc B* 370(1668):69–81
- Kaplan R, Friston KJ (2018) Planning and navigation as active inference. *Biol Cybern* 112(4):323–343
- Keramati M, Gutkin B (2014) Homeostatic reinforcement learning for integrating reward collection and physiological stability. *eLife* 3(e04811):1–26
- Khalsa SS, Adolphs R, Cameron OG, Critchley HD, Davenport JS, Feinstein JS, Feusner JD, Garfinkel SN, Lane RD, Mehling WE, Meuret AE, Nemeroff CB, Oppenheimer S, Petzschners FH, Pollatos O, Rhudy JL, Schramm LP, Simmons WK, Stein MB, Stephan KE, Van Den Bergh O, Van Diest I, von Leupoldt A, Paulus MP (2018) Interoception and mental health: a roadmap. *Biol Psychiatry Cogn Neurosci Neuroimaging* 3:501–513
- Kiebel SJ, Daunizeau J, Friston KJ (2008) A hierarchy of time-scales and the brain. *PLoS Comput Biol* 4(11):e1000209
- Kirchhoff MD, Froese T (2017) Where there is life there is mind: in support of a strong life-mind continuity thesis. *Entropy* 19(4):169
- Kirchhoff M, Parr T, Palacios E, Friston KJ, Kiverstein J (2018) The Markov blankets of life: autonomy, active inference and the free energy principle. *J R Soc Interface* 15(138):20170792
- Knight FH (1921) *Risk, uncertainty, and profit*. Sentry Press, New York
- Kozyreva A, Hertwig R (2019) The interpretation of uncertainty in ecological rationality. *Synthese*. <https://doi.org/10.1007/s11229-019-02140-w>
- Kräuchi K, Wirz-Justice A (1994) Circadian rhythm of heat production, heart rate, and skin and core temperature under unmasking conditions in men. *Am J Physiol* 267(3 Pt 2):R819–R829
- Krogh A, Lindhard J (1913) The regulation of respiration and circulation during the initial stages of muscular work. *J Physiol* 47:112–136
- Krupenye C, Kano F, Hirata S, Call J, Tomasello M (2016) Great apes anticipate that other individuals will act according to false beliefs. *Science* 354(6308):110–114
- Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A* 20(7):1434–1448
- Lee D, McGreevy BP, Barraclough DJ (2005) Learning and decision making in monkeys during a rock-paper-scissors game. *Cogn Brain Res* 25(2):416–430

- Levin M, Pezzulo G, Finkelstein JM (2017) Endogenous bioelectric signaling networks: exploiting voltage gradients for control of growth and form. *Annu Rev Biomed Eng* 19:353–387
- Levy I, Snell J, Nelson AJ, Rustichini A, Glimcher PW (2010) Neural representation of subjective value under risk and ambiguity. *J Neurophysiol* 103(2):1036–1047
- Lewis D (1973a) Causation. *J Philos* 70(17):556–567
- Lewis D (1973b) *Counterfactuals*. Basil Blackwell Ltd, Oxford
- Lewis D (1979) Counterfactual dependence and time's arrow. *Noûs* 13:455–476
- Limanowski J, Friston KJ (2018) 'Seeing the dark': grounding phenomenal transparency and opacity in precision estimation for active inference. *Front Psychol* 9:643
- Linson A, Clark A, Ramamoorthy S, Friston KJ (2018) The active inference approach to ecological perception: general information dynamics for natural and artificial embodied cognition. *Front Robot AI* 5:21
- Lyon P (2015) The cognitive cell: bacterial behavior reconsidered. *Front Microbiol* 6:264
- Lyon P (2019) Of what is "minimal cognition" the half-baked version? *Adapt Behav* 1–18
- Mackie GO, Burighel P (2005) The nervous system in adult tunicates: current research directions. *Can J Zool* 83:151–183
- Mathys CD, Lomakina EI, Daunizeau J, Iglesias S, Brodersen KH, Friston KJ, Stephan KE (2014) Uncertainty in perception and the hierarchical Gaussian filter. *Front Hum Neurosci* 8(825):1–24
- Maturana HR, Varela FJ (1980) *Autopoiesis and cognition: the realization of the living*. D. Reidel Publishing Company, Dordrecht
- McCoy JW (1977) Complexity in organic evolution. *J Theor Biol* 68(3):457–488
- McEwen BS, Stellar E (1993) Stress and the individual: mechanisms leading to disease. *Arch Intern Med* 153(18):2093–2101
- McGregor S, Baltieri M, Buckley CL (2015) A minimal active inference agent. [arXiv:1503.04187](https://arxiv.org/abs/1503.04187)
- Menaker M, Murphy ZC, Sellix MT (2013) Central control of peripheral circadian oscillators. *Curr Opin Neurobiol* 23(5):741–746
- Menon V, Uddin LQ (2010) Saliency, switching, attention and control: a network model of insula function. *Brain Struct Funct* 214(5–6):655–667
- Metzinger T (2017) The problem of mental action: Predictive control without sensory sheets. In: Metzinger T, Wiese W (eds) *Philosophy and predictive processing*, Chapter 19. MIND Group, Frankfurt am Main, pp 1–26
- Mikhalevich I, Powell R, Logan C (2017) Is behavioural flexibility evidence of cognitive complexity? How evolution can inform comparative cognition. *Interface Focus* 7(3):20160121
- Miracchi L (2019) A competence framework for artificial intelligence research. *Philos Psychol* 32(5):588–633
- Mirza MB, Adams RA, Mathys CD, Friston KJ (2016) Scene construction, visual foraging, and active inference. *Front Comput Neurosci* 10(56):1–16
- Mitchell A, Romano GH, Groisman B, Yona A, Dekel E, Kupiec M, Dahan O, Pilpel Y (2009) Adaptive prediction of environmental changes by microorganisms. *Nature* 460(7252):220–224
- Moore BR (2004) The evolution of learning. *Biol Rev* 79(2):301–335
- Moran RJ, Symmonds M, Dolan RJ, Friston KJ (2014) The brain ages optimally to model its environment: evidence from sensory learning over the adult lifespan. *PLoS Comput Biol* 10(1):e1003422
- Moreno A, Etxeberria A (2005) Agency in natural and artificial systems. *Artif Life* 11:161–175
- Morgan A (2018a) Mindless accuracy: on the ubiquity of content in nature. *Synthese* 195(12):5403–5429
- Morgan A (2018b) Pictures, plants, and propositions. *Mind Mach* 29(2):309–329
- Morville T, Friston KJ, Burdakov D, Siebner HR, Hulme OJ (2018) The homeostatic logic of reward. [bioRxiv](https://doi.org/10.1101/254444)
- Mugan U, MacIver MA (2019) The shift from life in water to life on land advantaged planning in visually-guided behavior. [bioRxiv](https://doi.org/10.1101/254444)
- Nakajima M, Imai K, Ito H, Nishiwaki T, Murayama Y, Iwasaki H, Oyama T, Kondo T (2005) Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science* 308(5720):414–415
- Neill WH (1979) Mechanisms of fish distribution in heterothermal environments. *Am Zool* 19(1):305–317
- Nicolis G, Prigogine I (1977) *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations*. Wiley, New York
- Nute D (1975) Counterfactuals. *Notre Dame J Formal Logic* 16(4):476–482

- Owens AP, Allen M, Ondobaka S, Friston KJ (2018) Interoceptive inference: from computational neuroscience to clinic. *Neurosci Biobehav Rev* 90:174–183
- Palacios ER, Razi A, Parr T, Kirchhoff MD, Friston KJ (2020) On Markov blankets and hierarchical self-organisation. *J Theor Biol* 486:110089
- Palmer CJ, Seth AK, Hohwy J (2015) The felt presence of other minds: predictive processing, counterfactual predictions, and mentalising in autism. *Conscious Cogn* 36:376–389
- Parr T, Friston KJ (2017) Uncertainty, epistemics and active inference. *J R Soc Interface* 14(20170376):1–10
- Parr T, Friston KJ (2018a) The anatomy of inference: generative models and brain structure. *Front Comput Neurosci* 12:90
- Parr T, Friston KJ (2018b) The discrete and continuous brain: from decisions to movement—and back again. *Neural Comput* 30:1–29
- Parr T, Corcoran AW, Friston KJ, Hohwy J (2019) Perceptual awareness and active inference. *Neurosci Conscious* 5(1):niz012
- Paulus MP, Stein MB (2006) An insular view of anxiety. *Biol Psychiat* 60(4):383–387
- Pavlov IP (1902) The work of the digestive glands. Charles Griffin & Co., Ltd, London
- Payzan-LeNestour E, Bossaerts P (2011) Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Comput Biol* 7(1):e1001048
- Pearl J (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Publishers, San Mateo
- Penny W, Stephan K (2014) A dynamic Bayesian model of homeostatic control. *Lect Notes Comput Sci* 8779:60–69
- Penny WD, Zeidman P, Burgess N (2013) Forward and backward inference in spatial cognition. *PLoS Comput Biol* 9(12):e1003383
- Perry CJ, Barron AB, Cheng K (2013) Invertebrate learning and cognition: relating phenomena to neural substrate. *Wiley Interdiscip Rev Cogn Sci* 4(5):561–582
- Peters A, McEwen BS, Friston KJ (2017) Uncertainty and stress: why it causes diseases and how it is mastered by the brain. *Prog Neurobiol* 156:164–188
- Petzschner FH, Weber LAE, Gard T, Stephan KE (2017) Computational psychosomatics and computational psychiatry: toward a joint framework for differential diagnosis. *Biol Psychiat* 82:421–430
- Pezzulo G (2008) Coordinating with the future: the anticipatory nature of representation. *Mind Mach* 18(2):179–225
- Pezzulo G (2014) Why do you fear the bogeyman? An embodied predictive coding model of perceptual inference. *Cogn Affect Behav Neurosci* 14(3):902–911
- Pezzulo G (2017) Tracing the roots of cognition in predictive processing. In: Metzinger T, Wiese W (eds) *Philosophy and predictive processing*, Chapter 20. MIND Group, Frankfurt am Main, pp 1–20
- Pezzulo G, Castelfranchi C (2007) The symbol detachment problem. *Cogn Process* 8(2):115–131
- Pezzulo G, Castelfranchi C (2009) Thinking as the control of imagination: a conceptual framework for goal-directed systems. *Psychol Res* 73(4):559–577
- Pezzulo G, Cisek P (2016) Navigating the affordance landscape: feedback control as a process model of behavior and cognition. *Trends Cogn Sci* 20(6):414–424
- Pezzulo G, Rigoli F, Friston KJ (2015) Active inference, homeostatic regulation and adaptive behavioural control. *Prog Neurobiol* 134:17–35
- Pezzulo G, Cartoni E, Rigoli F, Pio-Lopez L, Friston KJ (2016) Active inference, epistemic value, and vicarious trial and error. *Learn Memory* 23(7):322–338
- Pezzulo G, Kemere C, van der Meer MAA (2017) Internally generated hippocampal sequences as a vantage point to probe future-oriented cognition. *Ann N Y Acad Sci* 1396(1):144–165
- Pezzulo G, Rigoli F, Friston KJ (2018) Hierarchical active inference: a theory of motivated control. *Trends Cogn Sci* 22(4):294–306
- Powers WT (1973) Feedback: beyond behaviorism. *Science* 179(4071):351–356
- Preuschoff K, Quartz SR, Bossaerts P (2008) Human insula activation reflects risk prediction errors as well as risk. *J Neurosci* 28(11):2745–2752
- Quadt L, Critchley HD, Garfinkel SN (2018) The neurobiology of interoception in health and disease. *Ann N Y Acad Sci* 1428(1):112–128
- Raby CR, Alexis DM, Dickinson A, Clayton NS (2007) Planning for the future by western scrub-jays. *Nature* 445(7130):919–921
- Ramsay DS, Woods SC (2014) Clarifying the roles of homeostasis and allostasis in physiological regulation. *Psychol Rev* 121(2):225–247

- Ramsay DS, Woods SC (2016) Physiological regulation: how it really works. *Cell Metab* 24(3):361–364
- Ramstead MJD, Badcock PB, Friston KJ (2018) Answering Schrödinger's question: a free-energy formulation. *Phys Life Rev* 24:1–16
- Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2(1):79–87
- Read CR, Garnier S, Beekman M, Latty T (2015) Information integration and multiattribute decision making in non-neuronal organisms. *Anim Behav* 100:44–50
- Redish AD (2016) Vicarious trial and error. *Nat Rev Neurosci* 17(3):147–159
- Redshaw J, Bulley A (2018) Future-thinking in animals: Capacities and limits. In: Oettingen G, Sevincer AT, Gollwitzer PM (eds) *The psychology of thinking about the future*, Chapter 2. The Guilford Press, New York, pp 31–51
- Requin J, Brener J, Ring C (1991) Preparation for action. In: Jennings JR, Coles MGH (eds) *Handbook of cognitive psychophysiology: central and autonomic nervous system approaches*, chapter 4. Wiley, New York, pp 357–448
- Rust MJ, Markson JS, Lane WS, Fisher DS, O'Shea EK (2007) Ordered phosphorylation governs oscillation of a three-protein circadian clock. *Science* 318(5851):809–812
- Sales AC, Friston KJ, Jones MW, Pickering AE, Moran RJ (2019) Locus coeruleus tracking of prediction errors optimises cognitive flexibility: an active inference model. *PLoS Comput Biol* 15(1):e1006267
- Salman H, Libchaber A (2007) A concentration-dependent switch in the bacterial response to temperature. *Nat Cell Biol* 9(9):1098–1100
- Sanchez-Fibla M, Bernardet U, Wasserman E, Pelc T, Mintz M, Jackson JC, Pennartz CMA, Verschure PFMJ (2010) Allostatic control for robot behavior regulation: a comparative rodent-robot study. *Adv Compl Syst* 13(3):377–403
- Schacter DL, Addis DR (2007) The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philos Trans R Soc B Biol Sci* 362(1481):773–786
- Schrödinger E (1992) *What is life? With "Mind and matter" and "Autobiographical sketches"*. Cambridge University Press, Cambridge
- Schulkin J, Sterling P (2019) Allostasis: a brain-centered, predictive mode of physiological regulation. *Trends Neurosci* 42(10):740–752
- Schwartenbeck P, FitzGerald T, Dolan RJ, Friston KJ (2013) Exploration, novelty, surprise, and free energy minimization. *Front Psychol* 4(710):1–5
- Schwartenbeck P, FitzGerald THB, Mathys CD, Dolan R, Kronbichler M, Friston KJ (2015) Evidence for surprise minimization over value maximization in choice behavior. *Sci Rep* 5(16575):1–14
- Schwartenbeck P, Passecker J, Hauser TU, FitzGerald THB, Kronbichler M, Friston KJ (2019) Computational mechanisms of curiosity and goal-directed exploration. *eLife* 8:e41703
- Segundo-Ortin M, Calvo P (2019) Are plants cognitive? A reply to Adams. *Stud Hist Philos Sci* 73:64–71
- Seifert U (2012) Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep Prog Phys* 75(12):126001
- Sengupta B, Stemmler MB, Friston KJ (2013) Information and efficiency in the nervous system—a synthesis. *PLoS Comput Biol* 9(7):e1003157
- Seth AK (2013) Interoceptive inference, emotion, and the embodied self. *Trends Cogn Sci* 17(11):565–573
- Seth AK (2014) A predictive processing theory of sensorimotor contingencies: explaining the puzzle of perceptual presence and its absence in synesthesia. *Cogn Neurosci* 5(2):97–118
- Seth AK (2015) The cybernetic Bayesian brain: From interoceptive inference to sensorimotor contingencies. In: Metzinger T, Windt JM (eds) *Open MIND*. MIND Group, Frankfurt am Main, pp 1–24
- Seth AK, Friston KJ (2016) Active interoceptive inference and the emotional brain. *Philos Trans R Soc B* 371(1708):1–10
- Seth AK, Suzuki K, Critchley HD (2012) An interoceptive predictive coding model of conscious presence. *Front Psychol* 2(395):1–16
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423
- Shipp S (2016) Neural elements for predictive coding. *Front Psychol* 7(1792):1–21
- Shipp S, Adams RA, Friston KJ (2013) Reflections on agranular architecture: predictive coding in the motor cortex. *Trends Cogn Sci* 36(12):706–716
- Smith GP (2000) Pavlov and integrative physiology. *Am J Physiol Regul Integr Comp Physiol* 279(3):R743–R755

- Smith R, Thayer JF, Khalsa SS, Lane RD (2017) The hierarchical basis of neurovisceral integration. *Neurosci Biobehav Rev* 75:274–296
- Smith-Ferguson J, Beekman M (2019) Who needs a brain? Slime moulds, behavioural ecology and minimal cognition. *Adapt Behav*. <https://doi.org/10.1177/1059712319826537>
- Solway A, Botvinick MM (2012) Goal-directed decision making as probabilistic inference: a computational framework and potential neural correlates. *Psychol Rev* 119(1):120–154
- Spencer H (1867) *First principles*, 2nd edn. Williams & Norgate, London
- Spratling MW (2017) A review of predictive coding algorithms. *Brain Cogn* 112:92–97
- Sprigge TLS (1970) *Facts, words and beliefs*. Routledge & Keegan Paul, London
- Srinivasan MV, Laughlin SB, Dubs A (1982) Predictive coding: a fresh view of inhibition in the retina. *Proc R Soc B* 216(1205):427–459
- Stalnaker RC (1968) A theory of conditionals. In: Rescher N (ed) *Studies in logical theory*, American Philosophical Quarterly supplementary monograph series. Basil Blackwell Ltd, Oxford, pp 98–112
- Stanley ML, Stewart GW, De Brigard F (2017) Counterfactual plausibility and comparative similarity. *Cogn Sci* 41(Suppl 5):1216–1228
- Steiner AP, Redish AD (2014) Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. *Nat Neurosci* 17(7):995–1002
- Stephan KE, Manjaly ZM, Mathys CD, Weber LAE, Paliwal S, Gard T, Tittgemeyer M, Fleming SM, Haker H, Seth AK, Petzschner FH (2016) Allostatic self-efficacy: a metacognitive theory of dyshomeostasis-induced fatigue and depression. *Front Hum Neurosci* 10(550):1–27
- Sterelny K (2003) *Thought in a hostile world: the evolution of human cognition*. Blackwell Publishing, Malden
- Sterling P (2004) Principles of allostasis: optimal design, predictive regulation, pathophysiology and rational therapeutics. In: Schulkin J (ed) *Allostasis, homeostasis, and the costs of physiological adaptation*, Chapter 1. Cambridge University Press, Cambridge, pp 17–64
- Sterling P (2012) Allostasis: a model of predictive regulation. *Physiol Behav* 106(1):5–15
- Sterling P, Eyer J (1988) Allostasis: A new paradigm to explain arousal pathology. In: Fisher S, Reason J (eds) *Handbook of life stress, cognition and health*, Chapter 34. Wiley, New York, pp 629–649
- Suddendorf T, Corballis MC (1997) Mental time travel and the evolution of the human mind. *Genet Soc Gen Psychol Monogr* 123(2):133–167
- Suddendorf T, Corballis MC (2007) The evolution of foresight: what is mental time travel, and is it unique to humans? *Behav Brain Sci* 30(3):299–313
- Suddendorf T, Redshaw J (2017) Anticipation of future events. In: Vonk J, Shackelford TK (eds) *Encyclopedia of animal cognition and behavior*. Springer, Berlin
- Suddendorf T, Bulley A, Miloyan B (2018) Prospection and natural selection. *Curr Opin Behav Sci* 24:26–31
- Sweis BM, Thomas MJ, Redish AD (2018) Mice learn to avoid regret. *PLoS Biol* 16(6):e2005853
- Tagkopoulou I, Liu Y-C, Tavazoie S (2008) Predictive behavior within microbial genetic networks. *Science* 320(5881):1313–1317
- Tang SKY, Marshall WF (2018) Cell learning. *Curr Biol* 28(20):R1180–R1184
- Teff KL (2011) How neural mediation of anticipatory and compensatory insulin release helps us tolerate food. *Physiol Behav* 103(1):44–50
- Thompson E (2007) *Mind in life: biology, phenomenology and the sciences of mind*. Harvard University Press, Cambridge
- Todd W (1964) Counterfactual conditionals and the presuppositions of induction. *Philos Sci* 31(2):101–110
- Tschantz A, Seth AK, Buckley CL (2019) Learning action-oriented models through active inference. *bioRxiv*
- Van de Cruys S (2017) Affective value in the predictive mind. In: Metzinger T, Wiese W (eds) *Philosophy and predictive processing*, Chapter 24. MIND Group, Frankfurt am Main, pp 1–21
- van Duijn M, Keijzer F, Franken D (2006) Principles of minimal cognition: casting cognition as sensorimotor coordination. *Adapt Behav* 14(2):157–170
- Verschure PFMJ, Pennartz CMA, Pezzulo G (2014) The why, what, where, when and how of goal-directed choice: neuronal and computational principles. *Philos Trans R Soc B* 369(1655):20130483
- Vincent P, Parr T, Benrimoh D, Friston KJ (2019) With an eye on uncertainty: modelling pupillary responses to environmental volatility. *PLoS Comput Biol* 15(7):e1007126
- Wen Y, Zhou W, Zhu X, Cheng S, Xiao G, Li Y, Zhu Y, Wang Z, Wan C (2015) An investigation of circadian rhythm in *Escherichia coli*. *Biol Rhythm Res* 46(5):753–762

- Wheeler M (2011) Mind in life or life in mind? Making sense of deep continuity. *J Conscious Stud* 18(5):148–168
- Wiener N (1961) *Cybernetics: Or control and communication in the animal and the machine*, 2nd edn. MIT Press, Cambridge
- Wiese W (2017) Action is enabled by systematic misrepresentations. *Erkenntnis* 82(6):1233–1252
- Wiese W, Metzinger T (2017) Vanilla PP for philosophers: a primer on predictive processing. In: Metzinger T, Wiese W (eds) *Philosophy and predictive processing*, Chapter 1. MIND Group, Frankfurt am Main, pp 1–18
- Williams D (2018) Predictive minds and small-scale models: Kenneth Craik’s contribution to cognitive science. *Philos Explor* 21(2):245–263
- Williams D, Colling L (2018) From symbols to icons: the return of resemblance in the cognitive neuroscience revolution. *Synthese* 195(5):1941–1967
- Yon D, de Lange FP, Press C (2019) The predictive brain as a stubborn scientist. *Trends Cogn Sci* 23(1):6–8
- Zénon A, Solopchuk O, Pezzulo G (2018) An information-theoretic perspective on the costs of cognition. *Neuropsychologia* 123:5–18
- Zwicker D, Lubensky DK, ten Wolde PR (2010) Robust circadian clocks from coupled protein-modification and transcription-translation cycles. *Proc Natl Acad Sci* 107(52):22540–22545

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

4

Embodiment in mind: The role of rhythmic visceral dynamics in cognitive development

The previous two chapters have dealt with the formalities of homeostasis, allostasis, and cognition as understood from the perspective of the free energy principle. While I have tried to animate these discussions with empirical examples where possible, their conceptual nature has necessitated a fairly abstract treatment for the most part. However, with these conceptual foundations now in place, the remaining chapters can focus on more concrete analyses of cognitive and physiological dynamics.

The present chapter continues to interrogate the relation between cognition and biological regulation under active inference, this time from the perspective of *embodiment*. After providing a brief précis of the broad spectrum of philosophical views in the embodied cognition literature, I examine how the notion of embodiment has been extended to active inference (and interoceptive inference in particular). Although active inference is compatible with an assortment of claims about embodiment, it seems neither motivated by, nor committed to, any substantive philosophical account of embodied cognition. Nonetheless, I will argue that active inference does inspire valuable insights about the role of visceral dynamics in cognitive development.

Embodiment in mind: The role of rhythmic visceral dynamics in cognitive development

Andrew W. Corcoran, Kelsey Perrykkad, Daniel Feuerriegel,
Jonathan Robinson

Abstract

Predictive processing has attracted widespread interest from the scientific and philosophical community, with many championing its capacity to resolve longstanding debates about the embodiment of cognition. This article sets out to appraise the merits of such claims, focusing in particular on recent formulations of ‘embodied active inference’ under the free energy principle. Our analysis leads us to conclude that most of these accounts invoke weak, potentially trivial conceptions of embodiment; those that make stronger philosophical claims do so independently of the active inference framework. We suggest a more compelling account of embodied active inference can be motivated by adopting a diachronic perspective that addresses how rhythmic physiological activity shapes brain development *in utero*. We characterise three candidate mechanisms through which this ‘visceral afferent training’ hypothesis might be realised: Activity-dependent neuronal development, periodic signal modelling, and oscillatory network alignment.

Keywords: Embodied cognition; Active inference; Foetal development; Foetal heart rate; Cardiac cycle; Brain-body communication

4.1 Introduction

What is the relation between body and mind? This fundamental question remains a deeply divisive topic within modern cognitive science. Attempts to settle it have traditionally fallen into one of two camps: For dualists like Plato and Descartes, the corporeal and the mental are essentially distinct; the body is a mere vessel in which the mind (or soul) temporarily dwells. For monists like Aristotle and Spinoza, however, body and mind are inextricably bound – “one and the same thing” (Spinoza 2017, 3p2s).

Contemporary debates in philosophy, psychology, and neuroscience still bear the marks of this schism. Although orthodox *cognitivism* rejects the ontological commitments of the Platonic-Cartesian tradition, its quest to understand the mind in terms of abstract computational principles that govern internal representations of external states of affairs has long been accused of harbouring a tacit allegiance to dualism (see, e.g., Damasio 1994; Haugeland 1998; Noë 2009; Searle 1980; van Gelder 1995). One illustration of this legacy is the tendency to conceive of mental processes as a kind of software that could be run on a variety of different hardware systems – a view that renders the particularities of the physical system irrelevant for the purposes of cognitive theorising. Over the past three decades, however, many have contested that the mind *cannot* be properly understood without taking its physical embodiment and environmental embeddedness into account. Proponents of such views, which inherit from the Aristotelian-Spinozan tradition by way of pragmatism and phenomenology, march under the banner of *embodied cognition*.

While orthodox and embodied theories of cognition are typically construed as being at loggerheads with one another, recent advances in computational neuroscience have sparked hopes of reconciliation. In this paper, we evaluate the prospects of a harmonious resolution of the embodiment debate, focusing on conceptual and empirical developments within the predictive processing (or more specifically, *active inference*) literature. We analyse some of the ways in which embodiment has been commonly understood under this framework, and the contributions it may (or may not) be able to offer to broader philosophical debates about the nature of cognition. We argue that progress towards a properly embodied understanding of cognition may be attainable under active inference, but only after a fundamental shift in theoretical perspective.

To be clear, we do not set out to analyse the various reconciliatory claims made on behalf of active inference – nor indeed to assess the virtues of the latter as a general approach to the study of cognition (embodied or otherwise). Neither do we attempt to settle any broader philosophical disputes between traditional and embodied cognitivists of various stripes. Rather, our aim here is to clarify the explanatory role of embodiment within

the scheme of active inference (or certain influential expositions thereof), to evaluate the empirical evidence adduced in support of such claims, and to gesture towards a more substantive theory of the role of embodiment in the predictive mind.

This paper is structured as follows: Section 4.2 briefly surveys some of the most prominent themes within the history of embodied cognition, exploring various ways in which the embodiment thesis has been interpreted. Section 4.3 introduces active inference and some of the ways it has been argued to support embodied theories of cognition. We contend that much of the theoretical and empirical work advanced in the name of embodied active inference depends on a rather weak conception of embodiment, one which shares little of the revolutionary spirit animating its precursors. Section 4.4 argues that a more substantive version of embodiment requires the adoption of a diachronic perspective on brain-body interaction. To this end, we introduce the *visceral afferent training* hypothesis, which aims to give an account of how rhythmic physiological dynamics condition the development of the foetal brain. Section 4.5 evaluates the strength of this hypothesis in the context of disputes about the nature of the relation between bodily and cognitive processes. Section 4.6 returns to the broader question of the unificatory potential of the active inference framework. Section 4.7 concludes by briefly highlighting how the visceral afferent training hypothesis complements recent thinking in embodied active inference research.

4.2 Getting a grip on embodied cognition

The claim that ‘cognition is embodied’ has gained widespread traction in the mind sciences, yet its precise meaning is surprisingly difficult to pin down (Aizawa, 2015). Embodied cognition refers to a diverse assortment of views, not all of which are compatible with one another. To a first approximation, advocates of embodied cognition are committed to some form of *embodiment thesis*. Roughly (and minimally) expressed, this is the idea that the cognitive agent’s bodily constitution¹ “intrinsically constrains, regulates, and shapes the nature of [its] mental activity” (Foglia and Wilson 2013, p. 319). While a thoroughgoing examination of the various ways in which this thesis has been elaborated lies well beyond our scope, it will be useful to sketch out a spectrum along which embodiment theories are clustered according to the strength of their opposition against cognitivism (for further discussion, see Alsmith and de Vignemont 2012; Goldman and de Vignemont 2009; Miłkowski 2019; Shapiro 2019a; Wilson 2002).

¹As is common in this literature, our discussion of embodied states, processes, etc., pertains to bodily structures and activity beyond the central nervous system (or equivalent central processing architectures in the case of artificial systems). The assumption that mental states depend on central neural substrates (roughly, that animal cognition is ‘embrained’) is taken for granted.

4.2.1 Varieties of embodiment

Embodied cognition is perhaps most often associated with a provocative set of ideas that took root in the early 1990s. Seminal work by Varela, Thompson and Rosch (1991) and Brooks (1991) set the agenda for *strong* (or *radical*) interpretations of the embodiment thesis, which characteristically invoke a thoroughgoing rejection of (at least some of) the representationalist, computationalist, and internalist foundations on which the edifice of classical cognitivism was raised (e.g., Fodor 1975, 1981; Pylyshyn 1984). Strong forms of embodied cognition champion the radical dissolution of boundaries that are conventionally assumed to separate the brain, body, and world, ushering in a fundamental decentralisation and redistribution of the processes through which consciousness and cognition are produced (Chemero, 2009; Favela, 2014; Gallagher, 2017; Thompson and Varela, 2001).

Strong varieties of embodied cognition don't deny that the brain contributes to cognitive activity, but the role ascribed to it is dramatically reduced compared to the orthodox, 'neurocentric' view. Just as Brooks' 'subsumption architecture' engendered robots capable of autonomously navigating their environment without recourse to complex algorithms or central representations, biological organisms are said to couple with the world in ways that obviate the need for intensive computation over internal mental states. Rather, adaptive behaviour emerges by virtue of sensorimotor feedback loops that exploit the structured relations brought forth via agent-environment interactions (Barrett, 2011; Hutto and Myin, 2013; Wilson and Golonka, 2013). Insofar as the explanans availed by cognitivism fail to capture such phenomena, proponents of strong embodiment consider it incapable of delivering a complete cognitive science.

Not all subscribers to the embodiment thesis are so pessimistic, however. Those who endorse a more moderate position do not see embodied cognition as being intrinsically at odds with orthodox cognitivism, but consider them eminently compatible. Moderates claim that embodiment makes an important (perhaps unique) contribution to the cognitive economy at least some (but not necessarily all) of the time (Clark, 1997, 2008b). They sympathise with the radical's analysis of the complex interplay between the agent's physical constitution and its environmental milieu, but do not take this as warrant for abandoning explanations that appeal to (e.g.) representationally-rich internal computation. The move to a moderately embodied cognitive science is thus more of a reorientation than a revolution, one that augments the orthodox picture without substantially undermining its core theoretical commitments (Goldman, 2012).

Moderates are occasionally accused of purveying a rather *weak* or impoverished account of embodiment (e.g., Chemero 2013; Di Paolo 2018; Gallagher 2017). Standard-bearers

for embodied cognition do not generally aspire to weak versions of the embodiment thesis. In the extreme, weak embodiment amounts to a mere truism: Minds are implemented in physical systems, the precise nature of which plays some role in determining the kinds of states they can occupy. The range of colour sensations a particular organism might experience depends in part on the composition and organisation of photoreceptor cells in its retinae. Similarly, one's competence in various cognitive domains will generally degrade as one's blood alcohol concentration increases. Classical cognitivists don't deny that bodily states and structures may constrain or influence mental states and phenomenology – they simply dismiss such facts as incidental to the project of understanding the abstract nature of mind.

4.2.2 Causation versus constitution

One way of delineating weaker from more substantive notions of embodiment is to appeal to the distinction between causal and constitutive dependency relations (e.g., [Aizawa 2007](#); [Block 2005](#); [Prinz 2009](#)). To continue the previous example, normal brain function is causally dependent on blood chemistry: The composition of circulating blood must be kept within certain homeostatic bounds in order to enable normal brain activity to occur. The fact that deviations beyond such bounds impair cognitive function may be of interest to certain specialists; nonetheless, this does little to advance the case for a substantive interpretation of the embodiment thesis.² Causal factors that perturb or modulate the ordinary functioning of the cognitive system are differentiated here from the intrinsic properties of the system itself – that is, the cognitive operations or processes that are the object of cognitive scientific inquiry (indeed, the stability of cognitive function despite chemical fluctuations within homeostatic bounds suggests blood chemistry *per se* offers little insight into the workings of the mind – it is rather one of many background conditions that enable cognition to unfold in creatures like us).

Constitutive (rather than causal) accounts of embodiment argue that certain extra-neural structures are part-and-parcel of the representational currency and computational processes that make up the mind. Clark ([2008a](#); [2008b](#)) offers one such account, in which he claims that the supervenience base for mental states and cognitive operations (sometimes) extends beyond neural substrates into bodily mechanisms and external objects (cf. [Menary 2010a](#); [Wilson 1994](#)). Likewise, Noë's ([2004](#)) enactive account of visual perception eschews the trivial remark that what one sees depends on where one looks (i.e. a *merely* causal explanation), arguing instead that one's implicit knowledge or

²Enactivists like Shaun Gallagher ([2017](#)) counter that blood-borne molecules do indeed play a constitutive role in cognition by virtue of their contribution to affective experience, whereas others like Alva Noë ([2004](#)) would seem satisfied with a causal interpretation (see [Block 2005](#)). We will address some of the complex and controversial aspects of the causal-constitutive distinction in Section 4.5.

‘mastery’ of such sensorimotor contingencies is constitutive of perceptual experience (cf. [Gibson 1979](#); [Hurley 1998](#); [O’Regan and Noë 2001](#)). While the details of such accounts vary widely amongst theorists, the broader point here is that embodiment is construed as playing some important and distinctive role in the way cognition is realised – a role that is not susceptible to the sort of deflationary attitude with which the orthodox cognitivist dismisses weak conceptions of embodiment as accidental or trivial features of cognitive implementation.

Nevertheless, not all proponents of embodied cognition are committed to a strictly constitutive reading of the embodiment thesis. Goldman and de Vignement ([2009](#)) endorse an explicitly causal account of embodied representation in the domain of social cognition (but see [Goldman 2016](#)). Influential theories of *grounded cognition* ([Barsalou, 2008](#)) and *embodied simulation* ([Gallese and Sinigaglia, 2011](#)) might likewise be construed as offering causal accounts of the way bodies determine (or constrain) cognitive phenomena; although, as Shapiro ([2019b](#)) points out, the distinction between causal antecedents and constitutive elements may be murky (more on this later). For present purposes, it suffices to say that embodiment may be couched as playing a *non-trivial* causal role within the broader cognitive economy, although views of this sort tend to be more modest and conservative (i.e. more readily assimilated into the mainstream cognitivist orthodoxy) than most other species of embodied cognition.

4.3 Predictive processing and (embodied) active inference

Predictive processing theories of cognition have come to dominate the philosophical and scientific landscape over the past decade. Much like embodied cognition, predictive processing is something of an umbrella term capturing a range of positions that vary in their philosophical commitments and theoretical ambitions (for helpful overviews, see [Hohwy 2020a](#); [Wiese and Metzinger 2017](#)). Here, we focus our attention on *active inference* ([Friston et al., 2017a](#)), a computational framework developed under the *free energy principle* ([Friston et al., 2006](#); [Friston and Stephan, 2007](#)). Given the abundance of contemporary literature dealing with this framework, we limit ourselves here to a minimal exposition of key concepts.

Broadly speaking, active inference seeks to provide a formal explanation for the emergence of adaptive processes (e.g., action, perception, and learning) in complex biological systems such as ourselves. Under this scheme, adaptive dynamics are construed as emergent properties of self-organising systems that conform to a variational principle of least free energy, which states that biological systems must change in ways that decrease their free energy in order to survive ([Friston et al., 2006](#); [Friston, 2010](#)). Free energy

here is an information-theoretic quantity that places an upper bound on the negative log-probability (i.e. surprise or self-information) of an observation, given a (generative) model encoding beliefs about the way observations are generated.³ In the predictive processing literature, this quantity is often equated with *prediction error* (cf. predictive coding; [Huang and Rao 2011](#); [Rao and Ballard 1999](#)).

Two important corollaries of the free energy principle are that (1) free energy minimising systems infer the causes of their sensory inputs in an approximately Bayes-optimal fashion (cf. the Bayesian brain hypothesis; [Knill and Pouget 2004](#)), such that hidden states in the environment come to be internally represented in the form of probabilistic model parameters; and (2) free energy minimising systems will choose those actions most likely to realise expected sensory states ([Friston et al., 2012](#)). A formal description of the update rules underwriting these dynamics is provided by the process theory implementation of active inference ([Friston et al., 2017a](#)).

Despite its apparent computational, inferential, and representational commitments, there is no doubt that active inference departs from classical cognitivism on a number of fronts. One obvious example is the shift in emphasis from ‘bottom-up’ explanations of feature detection and scene (re)construction (e.g., [Marr 1982](#)) to ‘top-down’ inferences on sensory input. Perception on this view is fundamentally *pragmatic* and *action-oriented*, a subjective interpretation of the world conditioned by the agent’s beliefs and projects ([Clark, 2013, 2016](#); [Ramstead et al., 2020](#); [Williams, 2018a,b](#)). Moreover, the embedding of perceptual inference within interwoven cycles of action and perception, the dissolution of the boundaries separating perception, cognition, and action, and the infusion of cognitive processing with affective valence, exemplify just three of the ways in which active inference speaks to prominent themes within the broader embodied (embedded, extended, enactive, ecological, etc.) cognition literature.

It is not our intention to exhaustively explore such points of contact here, nor to make any definitive claims about the proper situation of the framework with respect to cognitivist versus embodied philosophies of mind (but see [Allen and Friston 2018](#); [Nave et al. 2020](#), for discussion). Neither do we dispute claims to the effect that active inference is ‘fundamentally embodied’ (e.g., [Allen and Friston 2018](#), p. 2460) – in part due to apparent widespread agreement on this point (even those who defend purportedly cognitivist interpretations of predictive processing grant some version of the embodiment thesis; e.g., [Hohwy 2016, 2018](#)), but mostly due to the fact that, as was hopefully brought into relief in Section 4.2, this epithet is essentially vacuous in the absence of further qualification. Our goals here, then, are to (1) interrogate how the concept of embodiment ought

³More technically, the generative model combines the likelihood of observing some data, given their causes (model parameters), and prior beliefs about those causes (specified as a probability distribution or density function on model parameters; [Friston 2009](#)).

to be understood in the active inference literature, (2) ask whether this concept fulfills any special function in the broader scheme of the framework, and (3) consider if active inference may deliver new insights about the nature of embodiment (and its attendant philosophical controversies).

4.3.1 Embodied models

An obvious place to begin is Friston’s extensive work on the free energy principle. The notion of embodiment is frequently invoked by Friston, most notably in connection with the concept of the generative model. Models are said to be embodied in cortical hierarchies, brains, and entire organisms (or more abstractly, their phenotypes; e.g., [Friston and Stephan 2007](#)). Biological systems, for instance, are supposed to “distil structural regularities from environmental fluctuations [...] and embody them in their form and internal dynamics,” such that they “become models of causal structure in their local environment” ([Friston et al. 2012](#), p. 2101). In an intriguing twist, the environment is also said to embody the agent, “in the sense that the physical states of the agent are part of the environment” ([Friston et al. 2011](#), p. 89). This formulation leads Friston to some interesting conclusions about the recursive implications of modelling (and being a model of) one’s environment, and the deeply existential ramifications of active inference more generally (cf. [Hohwy 2016, 2020b](#)).

Whether brains, organisms, and other biological systems really do embody models of their environments, or whether they merely lend themselves to being described as such, is a matter of contemporary debate (see, e.g., [Andrews 2021](#); [Baltieri et al. 2020](#); [van Es 2020](#)). Either way, the dual implication that constituents of the (literal or fictive) generative model (1) extend beyond the brain, and (2) are intimately bound up with (‘attuned’ to) environmental dynamics, would seem to comport well with strong interpretations of the embodiment thesis (see, e.g., [Bruineberg and Rietveld 2014](#); [Bruineberg et al. 2018](#)).

Nonetheless, the extent to which the embodiment of generative models speaks to the embodiment of mind is an open question. For instance, one might apply pressure on the implicit assumption that the ‘organism-level’ generative model is the appropriate target for cognitive scientific inquiry: If some partition of the model were found that delimited cognitive processes within neural substrates, the relevance of non-neural model components would require independent motivation. Importantly, however, active inference does not provide any principled means for locating such distinctions (cf. [Andrews 2021](#); [Clark 2017b](#); [Kirchhoff and Kiverstein 2019](#); [Ramstead et al. 2019](#)). Similarly, casting biological systems as generative models that are deeply attuned to their environments may

be consistent with strong-embodiment views (e.g., that agents enjoy direct perceptual access to environmental affordances), but does not *compel* them – additional philosophical work is required to demonstrate why such interpretations should be favoured over deflationary alternatives.⁴

4.3.2 Embodied feelings

A related line of embodied thought points to the role of the generative model’s physical instantiation in defining the conditions of the organism’s biological viability. The basic idea here is that cognition is embodied in the sense of being fundamentally geared towards the maintenance of homeostasis (Barrett, 2017; Damasio, 2018; Seth, 2015). The expected (homeostatic) states entailed by the agent’s phenotype thus constitute the normative criteria against which the agent’s current and future sensory states are evaluated (cf. Di Paolo 2005; Colombetti 2014; Thompson 2007). This idea opens the door to a deeply *affective* conception of mind, whereby bodily and emotional feeling states play a key role in guiding adaptive action (cf. Damasio 1994, 2010; Panksepp and Northoff 2009).

Several embodied interpretations of active inference have been developed on the premise that affective experience is rooted in the body’s physiological state. The first wave of such work provided a predictive coding style treatment of *interoception*, the (conscious or unconscious) sensory processing of internal bodily conditions (Pezzulo 2014; Seth 2013; for precursors with less explicit emphasis on embodiment, see Gu et al. 2013; Seth et al. 2012). Subsequent elaboration of these *interoceptive inference* accounts has sought to explain how aberrant interoceptive processing and autonomic regulation might engender various psychopathologies (Owens et al., 2018; Paulus et al., 2019; Quadt et al., 2018; Smith et al., 2020), including depression (Badcock et al., 2017; Barrett and Simmons, 2015; Barrett et al., 2016; Seth and Friston, 2016; Stephan et al., 2016) and anxiety/stress disorders (Clark et al., 2018; Gerrans and Murray, 2020; Linson et al., 2020; Peters et al., 2017).

The purported inseparability of cognitive and affective processing (Kiverstein and Miller, 2015; Pessoa, 2008), coupled with the deep link between physiological and emotional feeling states (Critchley and Garfinkel, 2017; Gu et al., 2019), make affect a prime target for embodied theories of active inference. It is noteworthy, then, that some active inference models of affective experience do not ascribe any special role to the body, relying

⁴Indeed, Friston’s characterisation of the agent’s attunement with (or recapitulation of) environmental properties seems equally at home with classical (computational-representational) interpretations of animal cognition (e.g., Gallistel 1989).

instead on domain-general computational principles pertaining to prediction error dynamics (Hesp et al., 2021; Joffily and Coricelli, 2013; Van de Cruys, 2017). While such models might ultimately prove compatible with their interoceptive inference counterparts (see Fernandez Velasco and Loev 2020), their existence implies that the active inference framework is not necessarily committed to an explicitly embodied account of affective experience.

4.3.3 Embodied selves

Perhaps the most significant development of active inference with respect to embodied cognition is not the putative impact of interoceptive states on affective experience, but rather the means by which interoceptive information is incorporated within the cognitive economy at large. Under active inference, interoceptive streams converge with exteroceptive (e.g., vision, audition) and proprioceptive (somatomotor) modalities at integrative regions of the cortical hierarchy (Gu et al., 2013; Pezzulo et al., 2015; Owens et al., 2018). Such multi- or cross-modal dynamics open the door for interoceptive information to influence a wide variety of cognitive domains, offering an principled explanation of the way internal bodily states may bias or condition perceptual decision-making (see, e.g., Allen et al. 2016; Pezzulo 2014; Pezzulo et al. 2018; cf. Barrett and Bar 2009).

Evidential support for the multimodal integration of bodily signals has accrued from studies investigating *bodily self-consciousness*, a fundamental constituent of self-awareness (Gallagher, 2000, 2005; Blanke and Metzinger, 2009). In experimental paradigms such as the rubber-hand (Botvinick and Cohen, 1998) and full-body illusions (Ehrsson, 2007; Lenggenhager et al., 2007; Mizumoto and Ishikawa, 2005), simultaneous visuo-tactile stimulation induces the uncanny experience of tactile sensations that seemingly arise from an artificial limb or body avatar (for discussion, see Apps and Tsakiris 2014; Aspell et al. 2012; Blanke 2012; Blanke et al. 2015; Limanowski and Blankenburg 2013). This effect is so compelling that subjects often report a sense of ownership over the alien limb or virtual body, as if it had been incorporated as part of (or in place of) their own body (Tsakiris 2010; cf. de Vignemont 2011).

Variants of these paradigms in which visual stimulation occurs in conjunction with periodic interoceptive events, such as heartbeats (Aspell et al. 2013; Heydrich et al. 2018; Suzuki et al. 2013; see also Sel et al. 2017) or breathing cycles (Adler et al., 2014; Allard et al., 2017; Betka et al., 2020; Monti et al., 2020), reveal that alterations in bodily self-consciousness can be induced via the integration of exteroceptive and interoceptive signals. That is, simply augmenting the visual representation of a virtual body(part)

with cardiac- or respiratory-synchronised pulsations is sufficient to modulate experiences of bodily ownership in the absence of concomitant physical stimulation of one’s body.

This body of work lends credence to the notion that the brain integrates multiple sources of sensory information to infer which parts of the environment are part of its body. Whether such inferences qualify as instances of embodied cognition is debatable; those of a cognitivist disposition may be inclined to argue that the mere representation of an object as part of oneself (or not) fails to motivate any claims about the embodiment of cognition. If anything, the directionality of these effects would seem to be precisely backwards from the perspective of embodiment: Such illusions showcase the susceptibility of body-related phenomenology to distortion by visual stimulation, rather than the capacity of bodily processes to influence one’s perception of the world.⁵

4.3.4 Embodied rhythms

Happily, we can eschew this concern by turning our attention to a complementary line of cross-modal research investigating the effect of bodily rhythms on cognition. Rather than presenting exteroceptive stimuli synchronously or asynchronously with a series of rhythmic interoceptive events such as the heartbeat (as in interoceptive variants of the rubber-hand and full-body illusions), cycle-timing paradigms investigate how sensorimotor and cognitive processing varies as a function of time *within* each physiological cycle. Of interest here are a subset of cycle-timing studies that involve emotionally neutral, non-body-related stimuli, which are immune to the worry that these phenomena may be domain-specific quirks of bodily/affective processing (for broader overviews of the literature, see [Azzalini et al. 2019](#); [Critchley and Garfinkel 2018](#)).

Most cycle-timing studies compare the difference between stimuli that are presented around the time of the heartbeat (more precisely, during the *systolic* time window in which the brain is receiving a burst of interoceptive input caused by the contraction of the heart), and stimuli presented between successive beats (i.e. during the *diastolic* period in which the heart is relaxing and refilling with blood). Early reports that sensorimotor processing is facilitated during diastole ([Birren et al., 1963](#); [Callaway and Layne, 1964](#); [Saari and Pappas, 1976](#); [Sandman et al., 1977](#); [Walker and Sandman, 1982](#)) have been replicated and extended in more recent years ([Edwards et al., 2007, 2008](#); [McIntyre et al., 2007, 2008](#); [Quelhas Martins et al., 2014](#); [Stewart et al., 2006](#); [Wilkinson et al., 2013](#); [Yang et al., 2017](#)). Furthermore, evidence that certain cognitive functions may be

⁵To adapt a more general argument from Clark ([2008a](#)), although the experience of bodily self-consciousness might require that some body is inserted within the sensorimotor loops mediating interactions between my brain and its environment, these studies indicate that it needn’t be my body in particular, nor indeed one very much like it.

enhanced (Fiacconi et al., 2016; Pramme et al., 2014, 2016; Rae et al., 2018), and certain forms of spontaneous action preferentially initiated (Galvez-Pol et al., 2020; Kunzendorf et al., 2019; Ohl et al., 2016) during systole has also accrued.

How do these findings (and analogous results from respiratory cycle-timing studies; Flexman 1974; Nakamura et al. 2018; Park et al. 2020; Perl et al. 2019; Waselius et al. 2019; Zelano et al. 2016) contribute to debates about embodied cognition? From one perspective, such data might seem to furnish incontrovertible evidence of the mind’s embodiment. After all, this literature attests to the pervasive influence of rhythmic internal dynamics on mental states, insofar as momentary fluctuations in afferent feedback determine whether an object enters into conscious awareness (in the case of near-threshold stimulation), when it is perceived (in the case of spontaneous action), and the extent to which it is subsequently processed and acted upon. And while the magnitude of such effects is admittedly modest, such fine margins may prove highly consequential in certain real-life situations (Azevedo et al. 2017; see also Fridman et al. 2019).

As with our earlier treatment of multisensory integration, however, the implications of such findings are not clear-cut. The cognitivist might be intrigued to learn that visceral organs exert subtle influences over perceptual decision-making in the visual domain, but needn’t concede that these effects are any more compelling than (say) the soporific effect of a heavy lunch. This rebuttal might seem too glib, but phasic fluctuations in perceptual acuity, alertness, reaction times, etc. are a well-established consequence of biological rhythms subtending several timescales (see, e.g., Jennings 1986), and yet are not generally adduced as compelling arguments in favour of embodied cognition. Instead, these phenomena belong to that class of factors whose influence upon cognitive processing and conscious experience is widely regarded as ‘merely’ causal.

What matters here, then, is not so much evidence that the heartbeat (breathing cycle, gut rhythm, etc.) modulates cognitive function, but rather the nature of the mechanism(s) underwriting such effects. These mechanisms are not fully understood, but converging evidence suggests that afferent discharge caused by systolic contraction exerts a transient inhibitory effect on cortical activity. As such, each incoming volley of cardio-afferent feedback induces a momentary perturbation on the brain’s ongoing dynamics, hindering its capacity to process information. An alternative (but not mutually exclusive) explanation cites the physical distention of blood vessels caused by the pulse wave as responsible for disturbing the sites where sensory information is registered, such as the retinae (Allen et al. 2019; see also Macefield 2003). Stimuli are simply more difficult to detect when they coincide with (and are obscured by) pulsatile motion (but see Grund et al. 2021).

Seen in this light, the cardio-afferent feedback thought to underlie the differences in sensorimotor and cognitive performance observed at different phases of the cardiac cycle appears to be a source of noise, a nuisance variable that impedes the brain’s capacity to go about its normal information-processing business. Although there are some lines of research that suggest the afferent signals associated with systolic contraction might benefit neural processing under certain contexts or conditions (e.g., [Garfinkel and Critchley 2016](#); [Pramme et al. 2016](#)), the tendency amongst active inference-inspired explanations of cycle-timing effects is to view the mechanical consequences of the heartbeat as a source of noise that lowers the precision over sensory feedback. In principle, one could envisage an agent whose cognitive capacities remain perfectly intact – indeed, are perhaps even improved on average – having eliminated the pulsatile ‘artifacts’ caused by the heartbeat (e.g., by replacing the heart with a continuous-flow device). Not only does the role of cardiac interoceptive feedback in consciousness and cognition appear to be more causal than constitutive in nature, its contribution seems more of a ‘bug’ than a ‘feature’.

To sum up: Active inference provides a powerful framework for explaining how the brain integrates multiple streams of information to construct conscious experience and regulate activity. Insofar as such schemes admit bodily states as important sources of sensory input, they would seem to open the door to a more embodied understanding of mind. But as we have highlighted, embodiment means different things to different people; while active inference might be hospitable to certain varieties of embodied cognition, the explicitly embodied versions of active inference analysed here do not compel strong interpretations of the embodiment thesis. Brains constantly track the evolving dynamics of their bodily states, and sometimes use this information to inform perception and action in other domains. Yet, on these accounts at least, it is the brain (and only the brain) that does the essential work of modelling the state of the body and the world. Visceral signals are accorded no special role within the inferential hierarchy, and appear to be treated no differently than other forms of sensory input (see also [Hohwy 2016](#); [Hohwy and Michael 2017](#)). Indeed, what appears on first blush to be some of the most promising evidence favouring the pervasive influence of interoceptive feedback on the mind turns out on closer inspection to be just another ‘noise trajectory’ ([Allen et al., 2019](#)) that the brain must learn to live with.

4.4 A diachronic perspective on embodiment

Thus far, we have evaluated the nature of embodiment under active inference from a broadly ‘synchronic’ perspective; that is, we have analysed how the active inference

framework might explain particular phenomena that occur over relatively brief time-frames within the mature cognitive agent. By contrast, relatively few researchers within the field have adopted an explicitly ‘diachronic’ or developmental perspective – although this situation is now beginning to change (e.g., [Atzil et al. 2018](#); [Ciaunica and Crucianelli 2019](#); [Ciaunica et al. 2021](#); [Fabry 2017a,b](#); [Fotopoulou and Tsakiris 2017](#); [Köster et al. 2020](#); [Martínez Quintero and De Jaegher 2020](#); [Montirosso and McGlone 2020](#); [Wozniak 2019](#)). The key idea we wish to explore in the remainder of this paper is that the most novel and interesting insights active inference has to offer the embodied cognition debate derive not from its account of how the adult brain balances converging streams of interoceptive and exteroceptive stimulation, but from the story it has to tell about the formative influence of visceral signals on the nascent mind. This idea highlights the critical role of embodiment in the formation of the mind – a role that may be obscured from view when considering the activity of mature cognitive systems.

The synchronic perspective encourages one to conceive of the cognitive agent as a fully-formed, pre-given entity, and attempts to analyse the structure of its behaviour under a given set of conditions in order to discover its internal logic. From this vantage, it is perfectly natural to view the brain as a central controller tasked with governing and coordinating various sorts of bodily activity, a view active inference inherits from mid-20th century cybernetics ([Seth 2015](#); cf. [Sperry 1952](#)). While this perspective is characteristic of traditional cognitivism, it is by no means unique to it – proponents of the embodiment thesis might likewise adopt a synchronic perspective when entering into disputes about the bounds of cognition, the aptness of representational and computational talk, and so on.⁶

By contrast, a diachronic approach seeks to examine how the cognitive system unfolds over time. While not necessarily undermining or contradicting conclusions derived from synchronic analysis, this approach may help to reframe or disrupt conventional assumptions about the nature of cognitive systems in ways that bring new insights into view. Recent focus on the embodied interactions between infants and caregivers, for instance, highlights how bodily and social exchanges during the early phases of childhood development play a fundamental role in realising homeostatic regulation, laying the foundations for the development of higher cognitive functions ([Fotopoulou and Tsakiris, 2017](#); [Seth and Tsakiris, 2018](#)). Analyses of this sort remind us that the brain undertakes a long and circuitous journey on its way to assuming its sovereign status, and that the sensory

⁶Note that our use of the term ‘synchronic’ is not meant to imply a static view of cognition. Cognitive dynamics clearly unfold over multiple timescales, thereby generating complex, nonlinear patterns of interaction. Rather, the thought is that the system consists of a relatively stable set of functional relations that evince predictable responses to different sorts of (external or self-generated) perturbation. More abstractly, one might think of a system that repeatedly visits a relatively small set of states, rather than one that periodically transitions to a new configuration that precludes the return to former states.

experiences encountered along this trajectory may have profound ramifications for the maturation of the cognitive system at large.

In what follows, we bring a diachronic perspective to bear on the emergence of so-called interoceptive noise trajectories and their regulation under active inference. The basic hypothesis we entertain is that the seemingly trivial (or perhaps mildly detrimental) fluctuations in neural activity associated with periodic physiological events such as the heartbeat are the residuum of a more fundamental developmental process. More specifically, oscillatory visceral inputs during the earliest stages of development are posited to play an instrumental role in sculpting the basic inferential architecture enabling the mature brain to effectively suppress or attenuate such perturbations. This idea, which we refer to as the *visceral afferent training* hypothesis, invites a more substantive interpretation of the contribution of rhythmic visceral dynamics in the formation of biological cognition, insofar as these signals may be responsible for inducing neuronal sensitivity to the ubiquitous regularities modelled in postnatal life.

4.4.1 Learning from within

Granting that the brain modulates its activity according to inferences about unfolding visceral dynamics, the question arises as to how the brain comes to model such processes in the first instance. One possibility is that the neurophysiological apparatus responsible for monitoring and controlling physiological variables is hard-wired by natural selection. Although the notion that brains come prepackaged with certain homeostatic prior expectations is fairly widespread within the active inference literature (e.g., [Allen and Friston 2018](#); [Allen and Tsakiris 2018](#); [Sims 2017](#)), the details concerning such innate programming are generally left unspecified. However, even if the foetal brain is ‘initialised’ with genetically-encoded prior expectations, the developmental and adaptive changes that (at least some) physiological parameters undergo over the lifespan suggests such priors can be modified or overwritten in response to prevailing environmental conditions (cf. [Yon et al. 2019](#)).

Generalising this point, it seems plausible that the basic structure of neurally-encoded generative models is inherited in the form of genetically-specified ‘wiring diagrams’, while the parameterisation of such models may depend on the physical stimulation to which their underlying constituents are exposed. In statistical parlance, this translates to the process of model fitting; under active inference, the optimisation of model parameters corresponds to learning ([Friston et al., 2016](#)). We conjecture that prenatal development encompasses both the formation of anatomical structures that support the hierarchical generative modelling of sensory states, and the tuning of model parameters

in response to the ‘training data’ availed by the internal and external (intrauterine) environments. While both of these processes clearly persist well beyond birth, we posit the foetal stage of prenatal development as a critical period for establishing basic forms of structure learning and predictive regulation that ground more sophisticated forms of model updating and selection in life outside the womb.

Our treatment here concentrates on the interoceptive data afforded by the cardiovascular system. This focus accords with the prominent theme of heart-brain interaction in both the recent interoceptive inference literature (surveyed in Section 4.3) and the broader psychophysiological tradition. This focus is also motivated by the fact that the heart is the first organ to form and begin functioning within the vertebrate embryo (Yutzey and Kirby, 2002). Even before the embryonic heart tube has acquired its familiar four-chambered morphology, it generates co-ordinated, rhythmic contractions resulting in electrical signals reminiscent of the adult electrocardiogram (Boullin and Morgan, 2005). Blood circulation is established by the fourth week of embryogenesis, while the neural tube is still closing. The circulatory system is thus fully operational before the nervous system has even begun to acquire its basic organisational structure (Stiles and Jernigan, 2010).

Contrary to the standard (synchronic) picture in which the mature brain orchestrates various autonomic activities according to systemic need, the embryonic brain is deeply ignorant of the physiological dynamics unfolding beyond its borders. As brainstem nuclei begin to self-organise, the developing brain likely acquires some capacity to detect physiological fluctuations arising from cardiovascular and other organ systems. Once receptive to perturbations generated by the body’s physiological processes, the task then is to infer the likely causes of such sensory inputs. Structurally, this situation is akin to standard predictive coding/active inference accounts of exteroceptive perception, whereby the brain is said to invert a generative model in order to infer the hidden causes of its sensory states. It is precisely this symmetry between interoceptive and exteroceptive modes of perceptual inference that prompts Hohwy and Michael (2017) to remark that there is ‘nothing special’ about the way bodily states affect sensory processing – they are just more hidden causes for the brain to infer.

It is not easy to pinpoint the precise stage at which the developing nervous system becomes receptive to information about the state of its environment. Nonetheless, it seems reasonable to posit the rhythmic pulsations of the circulatory system as among the first sensations registered by the nascent brain. Within the lower portion of the brainstem, the subnuclei of the *nucleus tracti solitarii* (NTS; a major viscerosensory relay centre for cardiovascular, respiratory, gastrointestinal, and gustatory input; Saper 2002) undergo an intensive period of cytoarchitectonic development during the final few

weeks of the first trimester (Cheng et al., 2006). Vagal nerve fibres linking this region to the aortic arch are already established by the time this period of cellular differentiation is underway (Cheng et al., 2004), with more widespread networks of innervation unfolding throughout the course of the second trimester.

Presumably, the kinds of inferences generated by neuronal populations at the brainstem level are relatively simple; neurons within the mature NTS may for example infer whether blood pressure is increasing or decreasing on the basis of baroreceptor activity, and likely do so without entertaining more complex hypotheses about the distal states of affairs driving such fluctuations. Even at this level, however, some rudimentary form of learning (belief updating) may be necessary in order to calibrate prior expectations over cardiovascular parameters. Foetal blood pressure increases linearly over the second half of pregnancy (Struijk et al., 2008), whereas heart rate decreases most rapidly between the 16th and 20th weeks of gestation, and more steadily thereafter (Pillai and James, 1990). If neuronal activity within the NTS encodes priors over these parameters, it is plausible that these neurons gradually adapt their response functions in line with accumulating sensory evidence of a sustained shift in such parameters (see Segar 1997, for discussion of such ‘chronic resetting’ during pre- and postnatal development).

The potential for more sophisticated modeling of cardio-afferent signals increases at higher levels of the interoceptive hierarchy, where longer epochs of cardiac activity can be extracted and integrated with other streams of sensory information. In the mature brain, NTS efferents project beyond the brainstem to a variety of higher neural centres (Loewy, 1981; Benarroch, 1993; Saper, 2002; Smith et al., 2017). Neuroanatomical studies indicate that key limbic components of this network emerge early in human ontogeny; the embryonic hypothalamus is discernible 5 weeks post-fertilisation, while the amygdaloid nuclei begin to differentiate between weeks 6 and 8 (Müller and O’Rahilly, 2006). If subcortical regions such as these are capable of receiving information from barosensitive brainstem circuits during the early postembryonic period, the periodic fluctuations of the cardiac rhythm could represent the first form of sensory patterning extracted by the integrative centres of the prenatal brain.

In sum, the early establishment of autonomous cardiovascular activity in the developing embryo, coupled with the emergence of brainstem and subcortical nuclei towards the end of the first trimester, provide the necessary substrates for basic interoceptive processing. The propagation of cardio-afferent information through brainstem structures to limbic regions marks the genesis of the brain’s capacity to “model its own dynamic noise trajectories” (Allen et al. 2019, p. 24), although at this early stage the visceral inputs that inscribe such trajectories are very much signals to be processed rather than noise to

be suppressed. This leads us to the hypothesis that rhythmic fluctuations in baroreceptor feedback constitute a salient ‘learning signal’ driving the adaptation of early brain networks and the generative models they entail. The most interesting upshot of this hypothesis is the possibility that these adaptive processes extend beyond the interoceptive domain to play a foundational role in the formation and structuring of the cognitive architecture at large. We explore this idea next.

4.4.2 Visceral afferent training drives activity-dependent neuronal development

Sensory stimulation has long been known to play a crucial role in the structural and functional development of the immature brain (e.g., [Wiesel and Hubel 1963a,b](#)). Even before visual and auditory pathways gain access to external stimuli, spontaneous bursts of neuronal activity are responsible for carving out precise patterns of network connectivity ([Friauf and Lohmann, 1999](#); [Hanganu-Opatz, 2010](#); [Kandler et al., 2009](#); [Katz and Shatz, 1996](#); [Mooney et al., 1996](#); [Penn and Shatz, 1999](#)). Spontaneous network activity has also been observed within various other regions of the central nervous system, including the spinal cord, brainstem, cerebellum, hippocampus, and neocortex ([Blankenship and Feller, 2010](#); [Feller, 1999](#); [Yuste, 1997](#)). It seems plausible that interoceptive input from early-developing peripheral organ systems could play a similarly instrumental role in fine-tuning the organisation of neural circuitry within nascent brainstem structures.

What’s interesting about this idea beyond the domain of interoceptive processing *per se* is the potential for correlated, temporally-structured patterns of evoked neural activity to influence the development of integrative centres beyond the brainstem. Given the extensive connectivity between hubs within the central autonomic network, and the various interfaces they share with other key brain regions, early exposure to periodic interoceptive stimulation could have profound effects on the way core brain networks are refined and remodelled – and by extension, on the way information is channeled through these structures.

From an active inference perspective, this notion suggests that the genetic specification of generative models might be relatively sparse, concentrating on the principles governing cellular differentiation and network formation. The imprecise, stereotypical circuitry laid down during these early stages of neurogenesis are subsequently refined and elaborated in accordance with Hebbian (and possibly other) principles ([Goodman and Shatz, 1993](#); [Kirkby et al., 2013](#); [Leighton and Lohmann, 2016](#)). The capacity for such remodelling (both in the physical sense of synaptic stabilisation and pruning, and in the more abstract sense of model parameterisation and reduction) essentially relieves the genome

of the burden of mapping out the precise developmental trajectory of specialised neural circuitry, a responsibility which is borne instead by domain-general learning mechanisms that remain operative throughout the lifespan.

Delegating the fate of neuronal wiring to activity-dependent adaptive processes is an efficient and flexible strategy for fine-tuning network connectivity. However, such an approach is only likely to succeed if the right kind of activity is reliably instantiated at the right stage of development. In the case of visual and auditory pathways, this problem is solved by inducing spontaneous activity on the basis of genetically programmed priors ([Katz and Shatz, 1996](#)). Such an arrangement may not be necessary within interoceptive pathways, however, on account of the early availability of viscerosensory input. As outlined above, the early development of the cardiovascular system in particular means that a continuous source of periodic afferent stimulation is already present when the NTS and higher centres come online. Since this input is essentially guaranteed in virtue of the fact that healthy brain development cannot proceed in the absence of a functional circulatory system, it constitutes a highly dependable stimulus for driving activity-dependent neuronal remodelling.

4.4.3 Visceral afferent training inculcates a model of periodic fluctuation

In the healthy foetus, average heart rate usually increases to a peak of ~ 170 beats per minute (bpm) at 9-10 weeks (i.e. around the same time NTS nuclei are undergoing an intensive period of maturation; [Cheng et al. 2006](#)), gradually decreasing thereafter ([Hornberger and Sahn, 2007](#); [Pillai and James, 1990](#)). This translates to a heartbeat every ~ 350 ms. Leaving aside the potential import of this stimulus as a driver of brain-stem network development, this signal presents perhaps the first opportunity for higher brain centres to model rhythmic activity originating from beyond the central nervous system. Hypothalamic and amygdaloid nuclei may for instance adapt to periodic activity conveyed via the NTS by establishing oscillatory network dynamics that synchronise to this input. Such phase-locked neural oscillations encode predictions about the timing of afferent stimuli, and may constitute the first step towards the development of more sophisticated neural models of external periodic events (including those necessary to quell physiological noise trajectories of the sort described by [Allen et al. 2019](#)).

In the adult brain, the amygdalae have been implicated in the generation of the heartbeat-evoked potential ([Park et al., 2018](#)), an electrophysiological response that may constitute a neural correlate of cardiac interoceptive prediction error ([Ainley et al., 2016](#); [Petzschner et al., 2019](#)). This observation lends credence to the idea that these nuclei

encode expectations about the periodic timing of baroreceptor activation. From a more general perspective, the perpetual reverberation of baroreceptor feedback through deep subcortical regions may be crucial for establishing the first neural representations of a stable pattern of variation – and thus, a generative process or hidden cause – in its external environment. While it is surely the case that all developing neuronal populations are subject to some form of external stimulation (even if the source of such stimulation originates from elsewhere within the brain), the crucial point here is that there exists some signal that can be identified as a recurrent pattern against the background hum of neural noise and metabolic activity. In carving out periodic baroreceptor feedback as a coherent and predictable occurrence amidst the tumult of the sensory flux, the brain takes its first step on a lifelong journey of ‘sense-making’, or resolving meaningful structure within one’s environment.

It is worth pausing to consider why, assuming the general shape of our story is on the right track, heartbeat-related afferent input ought to play such an important role in the emergence (more specifically, ‘training’ or ‘fitting’) of basic generative models. It is certainly plausible that other physiological processes give rise to periodic neural stimulation, especially if subcortical receptivity to such signals occurs later than we have assumed (e.g., the second half of pregnancy). However, cardiac feedback seems like a good candidate to focus on for two reasons: First, as mentioned above, the early development of the circulatory system guarantees such feedback is available as soon as the brain is sufficiently mature to detect it (which, moreover, raises the possibility of its active participation in the maturational process itself); second, the frequent and perpetual nature of such feedback renders it a more tractable learning signal than many other potential candidates, which occur less regularly and/or over slower timescales (e.g., changes in the chemical composition of the amniotic fluid or circulating blood due to maternal feeding).

Whether or not other periodic signals are easier to resolve and predict once cardio-afferent fluctuations have been captured within a subcortical generative model is unclear. Once in possession of a model of cardiac interoceptive dynamics, other kinds of periodic input might become easier for the brain to track. This is to say that brain networks may exploit (repurpose or generalise) an established model of cardiac dynamics to bootstrap the modelling of other periodic signals. Either way, once a model characterising the periodic nature of heartbeat-evoked neural activity has been established, the brain may begin to discern other patterns of sensory input in relation (or contrast) to this signature. In this sense, the heartbeat might inculcate an *ur*-concept of repetition or recurrence, in which the basic distinction between discrete states or phases is elevated to the recognition of a certain kind of patterned continuity – a (generative) process – unfolding over time.

Although the basic concept of repetition need not entail periodicity, the additional quality of regularity conferred by the heartbeat’s periodic nature is also notable. From a modelling perspective, the regularity of the cardiac cycle renders the timing of each afferent influx highly predictable, and thus easier to resolve as a recurring signal rather than a random sequence of unrelated events.⁷ Periodicity essentially confines the signal to a single, narrowband frequency channel that the target neural population can ‘tune in’ to; enabling the hidden source of this input to be easily separated (decorrelated) from competing background activity. A way of characterising this scenario in the language of active inference is to say that periodic signals have greater precision than their more irregular counterparts. As such, periodic stimuli generate salient sensory data that compel rapid model updates in line with sensory evidence. Once a model of the periodic process is established, evidence in favour of its parameterisation is rapidly accrued in virtue of its capacity to generate highly accurate predictions about the occurrence of precise sensory input.

4.4.4 Visceral afferent training signals promote self-organising brain dynamics

From a dynamical systems perspective, the periodic nature of the heartbeat might also serve a more mechanistic function as a pacemaker for oscillatory neural activity and information processing. The idea that the cardiac rhythm acts as a pacemaker for central and peripheral dynamics is not new (see, e.g., [Coleman 1921](#)), and has recently been revived in work linking the frequency architecture of neural and physiological systems ([Klimesch 2018](#); see also [Corcoran et al. 2018](#); [Tort et al. 2018](#)). Without digressing into these ideas too deeply, there is an obvious congeniality between such views and the empirical facts of early foetal development as sketched out above. For instance, [Klimesch \(2018\)](#) has proposed that many characteristic oscillatory dynamics observed in the brain and the body can be unified as part of a harmonically-arranged hierarchy that takes the heartbeat as its scaling factor. Ascribing the heartbeat as the fundamental rhythm from which all other members of the frequency hierarchy are derived makes perfect sense when one considers the early functional emergence of the cardiac system, and its physical influence on early brain development.

To be clear, we are not claiming that neural populations oscillate solely in response to cardiac or other sources of physiological activity. Neuronal oscillations are often

⁷Although the foetal heartbeat does evince variability, this characteristic feature of the cardiac time-series emerges gradually over the latter half of gestation ([DiPietro et al., 2015](#); [Visser et al., 1981](#); [Wheeler and Murrills, 1978](#)). It’s an interesting question from our perspective whether the increasing volatility of cardio-afferent feedback constitutes an additional source of complexity in the training of more advanced generative models.

described as spontaneous in character, and it seems plausible that developing neural networks may begin to evince oscillatory dynamics as random or intermittent firing patterns slowly converge towards attractor basins of synchrony ([Luhmann et al., 2016](#); [Thomason, 2018](#)). The point here, however, is that these emergent, self-organising dynamics may be guided or driven towards different regions of phase-space by external oscillatory input such as that generated by the cardiovascular system. By analogy to the way that the infant brain is equipped with the necessary neural machinery to acquire any natural language, but requires immersion within a particular linguistic environment in order to realise this linguistic potential, the thought here is that periodic visceral stimulation entrains a particular regime of oscillatory patterning (first in the brainstem, and then higher centres) amidst the brain’s earliest sensory experiences. While it might be possible in principle for maturing brain regions to settle into oscillatory regimes by dint of the intrinsic properties of their cellular constituents, and perhaps even to cycle through a repertoire of regimes as a consequence of inter-regional coupling relations that wax and wane over time, rhythmic input from the heart might establish a crucial point (or rather, orbit) of stability around which early subcortical activity organises itself.

The potential involvement of the cardiac rhythm as a pacemaker (or ‘order parameter’; see [Haken 1983](#)) in the self-organisation of neuronal oscillations is interesting from an embodied cognitive perspective because it shows yet another way in which visceral activity might shape emergent brain dynamics (cf. [Van Orden et al. 2012](#)). This notion is also relevant for active inference accounts that conceive of neuronal oscillations as a means by which predictions and prediction-errors are conveyed between neuronal populations (see, e.g., [Friston 2019b](#)). Specifically, the temporal structuring of neuronal oscillatory dynamics imposed by the heartbeat may help to organise network communication such that the transmission of information between different brain regions is facilitated or optimised (e.g., via the alignment of disparate oscillatory regimes in relation to a common fundamental frequency). If cardio-afferent feedback does play a role either in enabling oscillatory dynamics to emerge in the brain, or in determining the functional profile of those dynamics, then by extension, the cardiovascular system makes an important contribution to cognitive development.

But therein lies the rub: If any of the putative visceral afferent training mechanisms proposed in this section are to have any significant bearing on the embodiment debate, we need to show that these mechanisms are not ‘merely’ part of the background causal matrix on which brain development unfolds, but play a decisive role in defining this trajectory. This returns us to the distinction between causal and constitutive dependency relations raised in Section 4.2.2. Even if one is suspicious of the view that embodied accounts must be constitutive accounts on pain of triviality, the provision of yet another weakly-embodied causal account is unlikely to progress existing debates very far. Hence,

if active inference is to shed new light on the question of embodiment, it needs to motivate theoretical positions that are not susceptible to such deflationary rebuttals. We address this concern next.

4.5 Causation versus constitution redux

Let us be clear on the issue at hand. According to Adams and Aizawa ([Adams and Aizawa, 2008](#); [Aizawa, 2010](#)), there has been widespread conflation within the embodied cognition literature of things that affect cognitive processes and things that are (part of) cognitive processes. Failure to respect this distinction results in what they label the *coupling-constitution fallacy*, whereby components of the causal chain linking to (i.e. coupled with) some cognitive process are mistakenly classed as constituents (or realisers) of that process.⁸ As such, the burden is on the embodiment theorist to either (1) show that the cognitive process in question is indeed constitutively dependent on some bodily process, or (2) argue that certain species of causal dependency are sufficient for (non-trivial) embodiment, or (3) explain why the causal-constitutive distinction is unfit to adjudicate between cognitive and non-cognitive processes.

On first blush, the visceral afferent training hypothesis would seem to fall short of the constitutive standard stipulated by Adams and Aizawa. We have not claimed that the activity of developing visceral organ systems is intrinsically cognitive in any sense (indeed, we have not even asserted if and when foetal brain activity suffices for cognition); neither have we argued that visceral organ systems constitute an extension of the foetus' developing cognitive apparatus. By process of elimination, then, our hypothesis must turn on an essentially causal story, whereby visceral signals exert their influence on the brain by virtue of coupling relations. This being the case, our hypothesis would seem unlikely to furnish any substantive advances beyond existing accounts of embodied active inference – it simply speculates that cross-modal interactions between interoceptive and exteroceptive processing streams may be rooted in foetal development.

4.5.1 Constitution through causation?

This conclusion is too hasty, however. One reason for caution is that Adams and Aizawa's coupling-constitution fallacy is predicated on an essentially synchronic view of cognition ([Menary, 2010b](#)), and may as such be ill-suited to the diachronic perspective on offer here (see also [Kirchhoff 2015](#)). Moreover, hard-and-fast distinctions between

⁸Adams and Aizawa (2008) insist this distinction holds irrespective of whether the coupled processes in question form part of an integrated system (such as a biological organism).

causation and constitution become challenging in the context of temporally-extended processes (Shapiro, 2019a,b), especially when those processes are recursive and/or embedded within complex dynamical systems (Clark, 2008b; Gallagher, 2017; Kirchhoff, 2015). At least some causal factors appear to be ‘proper constituents’ of the processes they interact with, such that those processes simply couldn’t exist (and perhaps couldn’t be conceived) without them.⁹

Proper treatment of these metaphysical issues lies well beyond the scope of this paper; however, we take such objections as provisional justification for the notion that some cases of causal dependency may be deeply and irrevocably enmeshed in the generation of cognition – at least (or perhaps *especially*) when dealing with complex, nonstationary processes of the sort encountered in ontogenesis. From this perspective, visceral afferent training signals might be viewed as playing an instrumental role in driving the development of a particular kind of cognitive architecture, ‘configuring’ or ‘formatting’ the brain’s generative model such that it supports a particular range of inferences (e.g., about periodic structure embedded in sensory flows) and cognitive operations (e.g., the ability to distinguish internally- from externally-generated sensations). If this is right, and the causal ‘inscriptions’ of interoceptive inputs help organise the neural dynamics that ultimately constitute cognitive processing, such signals represent a crucial factor in determining the kinds of minds instantiated in animal brains.

A strong reading of the visceral afferent training hypothesis implies that the foundations of mind are laid down and moulded in accordance with multidimensional vectors of sensory information availed by the internal environment. Expressed differently: Bodies like ours – or more specifically, the temporally-structured physiological dynamics they entail – are necessary conditions for the emergence of minds like ours. This interpretation is reminiscent of (and perhaps continuous with) an early strain of thought in embodied cognition, whereby specific features of one’s morphology shape and constrain the sorts of concepts one can access or form (e.g., Barsalou 1999; Lakoff and Johnson 1980, 1999; cf. Shapiro 2019b). The idea that one’s entire conceptual apparatus might derive from a base set of concepts grounded in one’s physical interactions with the world shares a certain resemblance to the idea that the brain’s transactions with its visceral states cause proximal changes (i.e. in the format of the generative model) that propagate through to more distal, abstract mental representations (e.g., the elaboration of a model of oneself from more basic models).

⁹An example from Shapiro (2019a; 2019b): Sunlight is both a cause and a ‘proper constituent’ of photosynthesis, while frost is an incidental factor that may causally impinge on the leaf’s capacity to photosynthesise. The key idea is that some causal antecedents figure as ineliminable parts of the process at hand, not mere supports, adjuncts or modulators.

This point of contact with embodied theories of conceptual grounding is useful insofar as it reminds us that, as mentioned in Section 4.2.2, such theories can be read as rather weak interpretations of the embodiment thesis. Indeed, the indirect causal link we posit between visceral and cognitive processes might render the visceral afferent training hypothesis susceptible to a similar charge. Moreover, the mediated nature of this causal link belies another subtle departure from these earlier views: What really matters for our hypothesis is not so much the nature of one’s body, but rather that of the neural signals conveyed to the brain – signals that could in principle be instantiated in various organisms consisting of very different physiological and morphological features. Although manipulating the temporal characteristics of foetal heartbeat dynamics might be expected to have profound ramifications for the development of brain dynamics under our hypothesis, we have provided no reason to think that substituting the heart with a device that periodically stimulates the arterial baroreceptors according to age-matched parameters should make any notable difference.

4.5.2 Embodied mind or envatted brain?

Hypothetical considerations of this nature almost inevitably lead one to the perennial question of envatment, perhaps the acme of anti-embodiment thought experiments. In the classic ‘brain-in-a-vat’ scenario, the disembodied brain floats in a vat of life-sustaining nutrients, while communicating with a sophisticated computer by means of wires or transceivers affixed to various nerve endings. The intuition – at least for the cognitivist – is that this set up would be sufficient to replicate the cognitive processes and conscious experiences instantiated in an embodied brain, thus demonstrating that the mind is only contingently (i.e. causally) related to its non-neural body (e.g., [Metzinger 2003](#)). Embodiment theorists might counter that this scenario surreptitiously reintroduces a body and a world via the vat and computer simulation (e.g., [Hurley 2010](#); [Thompson and Cosmelli 2011](#); see also [Hohwy 2017](#)); however, unless a strong argument for the body’s unique involvement in the ‘preprocessing’ or ‘formatting’ of afferent input prior to central processing can be successfully mounted, it remains difficult to see how this position could support anything beyond a weak interpretation of the embodiment thesis.

We do not pursue the possibility of peripheral information processing and/or formatting here, since we are unaware of any resources within the active inference framework that would help arbitrate this question. Neither do we believe that a diachronic re-imagining of envatment can help defeat the argument: While establishing brain-computer interfaces that keep up with neurogenesis might prove technically challenging, we see no reason in principle why a ‘developmental’ programme of stimulation shouldn’t run just as well as its ‘mature’ counterpart. Indeed, it might even be the case that a diachronic variant of

the brain-in-a-vat would be capable of producing an even more convincing simulation of phenomenal experience, since there would be no potential for an uncanny disconnect between one’s intended actions and their expected consequences (consider how laggy or jittery mouse cursor movement disrupts the fluency of one’s actions and degrades one’s sense of agency in the virtual domain) – more specifically, there would be no opportunity for any such slippage to occur, since the developing brain is calibrated to the temporal structure of the computer-mediated action-perception loop from the get-go.

There is perhaps one unique possibility introduced by the visceral afferent training hypothesis that might argue against the perfect emulation of experience through envatment. This is the possibility that normal cognitive development (and mature functioning) is influenced not only by the sensory information generated by peripheral oscillators such as the heart, but also by the mechanical influence such oscillators exert over neural tissue. If this were the case, the envatted brain would require a special apparatus designed to mimic the periodic physical distention that would have been caused by a real heartbeat. Although this arrangement doesn’t entail the necessity of a cardiovascular system *per se* for normal cognition to arise, the necessity of some additional mechanism designed to emulate the functional profile of such bodily rhythms begins to seriously undermine the supposedly disembodied status of the envatted brain (cf. [Thompson and Cosmelli 2011](#)).

Setting this possibility aside, should the (nomo)logical possibility of a brain-in-a-vat scenario under the visceral afferent training hypothesis raise a red flag to defenders or advocates of embodied cognition (to whom the very concept of envatment is anathema)? Our view is that anyone who endorses active inference must be prepared to entertain such possibilities; as an essentially functionalist framework ([Colombo and Wright, 2018](#); [Hohwy, 2016](#)), active inference accommodates envatted brains as one of manifold potential solutions to the problem of free energy minimisation.¹⁰ As such, the brain-in-a-vat occupies one end of a spectrum of embodied active inference schemes, while cognitive systems extending beyond the body (or indeed, incorporating multiple bodies) populate the other – such is the versatility of the active inference formalism.

¹⁰On reflection, this realisation should perhaps come as little surprise: Proponents of embodied active inference are indeed rather fond of characterising perception as a kind of “controlled hallucination” (e.g., [Clark 2016](#); [Seth 2016](#)), a slogan which seems entirely in keeping with the Cartesian scepticism behind the brain-in-a-vat thought experiment (see, e.g., [Putnam 1981](#); see also [Hohwy 2016, 2017](#)). Clark for his part is prepared to admit such scepticism under embodied active inference, but dismisses it as “a mere distraction (a red herring)” ([2017a](#), p. 735).

4.6 Prospects for a unified philosophy of the embodied mind

Philosophers and cognitive scientists on both sides of the embodiment debate continue to claim active inference as a vindication of their views, in spite of their (supposedly fundamental) disagreements about the nature of cognition. We have suggested that the formal framework availed by active inference may accommodate a wide variety of philosophical positions on the question of embodiment without necessarily helping to arbitrate the merits and shortcomings of these alternatives. Perhaps active inference is silent on many of the details on which such arguments turn; perhaps the framework will eventually reach a level of maturity enabling it to rule out (or severely undermine) many of them. Presently, however, the promise of any unification, fusion, or synthesis of these divergent strands of thought remains largely unfulfilled; indeed, most theorists seem more interested in demonstrating that their favoured interpretation of the embodiment thesis fits within the active inference framework, rather than engaging in any genuine dialectical exchange with rival views.

But perhaps this conclusion presents an unduly pessimistic appraisal of the literature. A more optimistic interpretation might appeal to the following line of reasoning: By drawing together a large range of embodied theories under a single framework, active inference has helped bring previously unseen (or at least, neglected) points of contact between supposedly divergent views to the fore. For example, the ability to reformulate seemingly internalist, inferential, and representation-laden notions of generative modelling and prediction error minimisation in the language of ‘optimal grip’ and ‘attunement’ shows how the active inference framework affords both a common ground and a shared vocabulary – the foundations of any productive dialogue. Perhaps then there is reason to be hopeful that the active inference perspective might resolve (or dissolve) long-standing disputes between competing factions. Or, if it doesn’t help theorists to work out their differences, it might at least help them to work out which differences really *matter*.

As foreshadowed in the introduction, we did not set out to resolve any of the tensions amongst competing visions of embodied active inference, nor indeed the embodied cognition literature at large. What we have tried to do, rather, is bring some of these tensions to the fore. On our analysis, much of the empirical work conducted under the rubric of embodied active inference lends itself to a fairly deflationary interpretation; indeed, proponents of a more ‘radically’ embodied cognitive science would likely consider it entirely of a piece with mainstream cognitivism. This is not to say that embodied active inference ought to strive towards the more radical extreme of the embodiment

spectrum – it is simply to point out that most of this work has little in common with the strong aspirations and provocative rhetoric historically leveraged in the name of embodied cognition.

How does the visceral afferent training hypothesis fit within this picture? What we have attempted to do here is pursue a seemingly deflationary interpretation of the impact of embodiment on cognition (i.e. ‘visceral-feedback-as-noise-trajectory’) to mount a more substantive account of the relation between body and mind. While our account does not purport to radically overturn cognitivist-friendly interpretations of active inference, it goes beyond existing causal accounts of the way cognitive processing incorporates inferences about bodily states (i.e. interoceptive inference) and the sensory consequences of their temporal evolution (e.g., cycle-timing effects). Our diachronic account does not undermine synchronic analyses of the causal interaction between fully-fledged bodily and cognitive systems; rather, it tells a complementary story about the role interoceptive dynamics play in the emergence of the cognitive architecture itself.

4.7 Conclusion: The body as first teacher

The notion of the ‘first prior’ has become an important motif in recent work at the nexus of human development and embodied active inference ([Allen and Tsakiris, 2018](#); [Ciaunica et al., 2021](#)). This first prior might be construed as the phenotypic information bound up in one’s genetic inheritance, or more abstractly, as beliefs about one’s own existence and the imperative to realise sensory states that are conducive to that existence. The visceral afferent training hypothesis complements this idea with a more explicitly adaptive twist: The body is not only the organism’s first prior, but also its first *teacher* – one that provides reliable interoceptive instruction to guide the elaboration of the generative model throughout the prenatal period and beyond. And just like any good teacher, the body equips its student with precisely the right foundations to go out into the world, apply its knowledge to new problems, and acquire deeper levels of understanding along the way – thereby becoming a master in its own right.

5

Be still my heart: Cardiac regulation as a mode of uncertainty reduction

Having spent the past three chapters delving into various related theoretical and conceptual concerns, this point in the thesis marks the transition to a more empirical mode. The following three chapters report findings from empirical studies that aim to build on and complement the theoretical material introduced thus far. Given that empirical scientific writing general calls for a rather sparing, tightly-focused narrative, and given that the topics spanned by these studies are rather diverse, the manuscripts at the core of these chapters are preceded by lengthier passages of framing text than has hitherto been required. This text aims to provide additional contextualising information designed to complement that which is provided in the corresponding manuscript. It also serves to highlight points of contact with the content of other chapters, as well as broader continuities with the overarching themes of the thesis.

The shift in emphasis from a theoretical (but empirically-informed) to an empirical (but theoretically-inspired) mode is accompanied by an analogous shift in focus: While the past three chapters have been predominantly concerned with the nature of biological regulation, the next three chapters are progressively more concerned with the nature of cognitive regulation. To be sure, allostasis still plays an important conceptual role in the work that follows, but recedes further into the background as factors such as

attention and executive control come to the fore. (Parenthetically, there is something rather befitting about this arrangement; allostasis – or physiological regulation more generally – is something that tends to operate ‘behind the scenes’ for the most part, a deeply hidden cause in most cognitive scientific research.)

The thread that unites these chapters with what came before is the ubiquitous topic of uncertainty. Setting aside my earlier concern with the nature of the relation between uncertainty and biological (or cognitive) agents in general, the following experiments focus on the way a particular kind of cognitive architecture (i.e. that of the healthy human adult) deals with the demands of specific uncertainty-inducing scenarios. Attention is a natural focus for these studies, given its centrality within active inference as the mediator between perception and action, and its close formal relation to uncertainty via the modulation of (inverse) precision. In particular, the active inference conception of attention as covert action has an important role to play in each of the following chapters, serving to bridge earlier discussions about adaptive physiological and behavioural action with notions of mental action and cognitive control.

To briefly presage the content that follows, this chapter and the next continue the theme of heart-brain communication introduced in Chapter 4, dealing with the relation between allostatic (cardiac) action on the one hand, and perceptual dynamics (Chapter 5) or motor and cognitive control (Chapter 6) on the other. Chapter 7, by contrast, focuses more specifically on the relation between oscillatory brain dynamics and perception in the context of a paradigmatically cognitive domain: Language. Furthermore, each of the experiments reported in these chapters adopt psychophysiological methodologies, thereby entering into dialogue (in a more or less direct fashion) not only with the experimental literature discussed in previous chapters (most notably, Chapter 4), but also with a rich scientific tradition dating back at least as far as the pioneering work of Ivan Pavlov – arguably, the grandfather of allostatic research.

5.0.1 Psychophysiology: The science of embodiment

Psychophysiology has been characterised as a scientific mode of inquiry that seeks to “describe the mechanisms which *translate* between psychological and physiological systems of the organism” (Ax 1964, p. 8, emphasis in original). Determining the mapping between psychological and physiological phenomena is of course no trivial matter; if it were, the mind-body problem discussed in Chapter 4 would surely have been solved (or at least, reformulated in rather different terms) some decades ago. Nevertheless, the notion of a ‘translator’ that bridges the philosophically-treacherous waters between the physical (physiological) and the mental is suggestive of a bidirectional mode of communication or

interface that transcends weaker notions of ‘mere correlation’. Psychophysiology would thus seem to represent a scientific enterprise that is very much in allegiance with the spirit of embodied cognition – an impression reinforced by various entries in the Society for Psychophysiological Research’s back catalogue of *Presidential addresses* (see, e.g., [Jennings 1992](#)).

Some of the arguments presented in Chapter 4 may be read as expressing a certain degree of scepticism about the capacity of psychophysiological research to offer meaningful (let alone decisive) contributions to debates about the embodiment of mind. These arguments were focused on recent empirical work that has either drawn on theoretical elements of active inference to motivate hypotheses and/or interpret results, or generated evidence that has been adduced in support of embodied active inference. I suspect, however, that these arguments are apt to generalise beyond the active inference literature; the basic concern being that psychophysiological data are simply unable to arbitrate core philosophical disputes about substantive interpretations of the embodiment thesis. This – rather than widespread ignorance of the literature – would explain why psychophysiological evidence appears to have played no significant role in philosophical debates about embodiment over the past 30 years.

Granting the neutrality of psychophysiological research on philosophical matters of embodiment need not damage its capacity to shed light on the empirical nature of mind-body integration. This point notwithstanding, the difficulties inherent in mapping the translation from the physiological to the psychological domain should not be understated; after all, minds and bodies are extraordinarily complicated systems in their own right, and drawing strong inferences about the workings of one from those of the other demands a great deal of careful thought and rigorous experimentation ([Cacioppo et al., 2007](#)). Undoubtedly, researchers have on occasion been prone to elide important distinctions between different kinds of psychophysiological relation (e.g., correlational, mediational, regulatory; see [Jennings 1986](#); [Porges 1992](#)) – or, worse still, to mistake tenuous patterns of physiological fluctuation for reliable indicators of psychological constructs ([Obrist, 1981](#)). Nevertheless, a great deal of progress has been made towards the ‘decoding’ of psychophysiological phenomena since the field began to coalesce in the 1950s and ’60s.

This is certainly not the venue for a historical overview of the achievements of the field; however, one cannot help but remark in passing the impressive ingenuity, sophistication, and prescience of many ‘classic’ lines of psychophysiological research. While such work has remained an influential touchstone within the field, it has to some extent been rediscovered and reinvigorated of late by a broader cognitive (neuro)scientific community – one that has only recently been awakened to the pervasive impact of visceral signals on

neural dynamics and subjective experience (see, e.g., [Azzalini et al. 2019](#); [Critchley and Garfinkel 2018](#); [Park and Tallon-Baudry 2014](#); cf. [Cameron 2002](#); [Craig 2002](#); [Damasio 1996](#)). Exemplars of such work from the domain of cardiac psychophysiology will be introduced in this chapter and the next.

5.0.2 Psychophysiology and the study of covert processes

Arguably, all cognitive scientific inquiry is ultimately engaged in the investigation of covert activity, by virtue of the intrinsically unobservable nature of mental phenomena. The mind is the hidden cause *par excellence*; even one's own subjective experience of mentality is in some sense indirect, incomplete, and inferential ([Metzinger, 2003](#)). Psychophysiology is perhaps unique amongst the cognitive sciences insofar as it aims to learn something about the hidden structure of the mind through the study of other covert processes – albeit, those that are amenable to objective measurement.

One might legitimately wonder what the advantage of this arrangement is, given the potential difficulty of observing covert rather than overt processes (i.e. explicit behaviour), as well as the attendant complexities and pitfalls of interpreting such observations. A quick answer might be something to the tune of ‘the more data one can get, the better’ – given the apparent lack of any singular ‘royal road’ to the mind, it might be advisable to pursue any and every available avenue. Similarly pragmatic is the remark that not all cognitive processes lend themselves to analysis via objective behavioural (or subjective report) data – covert physiological responses might be the only clues one has to go on, or the only way to obtain information about a particular process without unduly disturbing it.

The reality is usually somewhere between these poles of abundance and paucity. Psychophysiological measures typically complement those garnered from overt behaviour and subjective experience, and are often interpreted in conjunction with them. For instance, recording heart rate fluctuations during the performance of a signalled reaction time task furnishes information about preparatory activity in the lead up to overt actions, a process that would be extremely difficult to tap with any precision via behavioural or subjective methods. Moreover, the psychophysiological information obtained during the preparatory interval can be analysed in conjunction with behavioural data to derive insights about the relation between covert and overt processes. Evidence that reaction time covaries with heart rate would indicate that some aspect of preparatory processing is indexed by cardiac modulation – perhaps heart rate changes are a manifestation of the systemic physiological adaptations that accompany enhanced sensory sensitivity; perhaps heart rate directly impinges upon the commission of a speeded response.

As alluded to above, alighting on the best explanation of such phenomena is often the most challenging aspect of psychophysiological research. The early decades of cardiac psychophysiology were dominated by high-profile disputes about the sensitivity of heart rate dynamics to sensory properties and task demands, and especially about the functional import of such dynamics (e.g., [Lacey and Lacey 1974](#); [Obrist 1981](#); for discussion, see Chapter 6). What became increasingly apparent over this period was that changes in cardiac activity associated with attention and information processing cannot simply be ascribed to the metabolic requirements induced by the task, nor to generic fluctuations in physiological arousal or inhibition. This body of evidence favours the view (promulgated perhaps most forcefully by John and Beatrice Lacey; [1974](#); [1978](#)) that cardiac adaptations are functionally relevant to the mental processes they accompany, not incidental consequences of them.

Casting such physiological dynamics in terms of covert action is a helpful reminder of the basic allostatic insight that, the existence of local homeostatic feedback mechanisms notwithstanding, physiological processes are regulated and co-ordinated by top-down mechanisms. Indeed, historically speaking, (psycho)physiologists have tended to focus almost exclusively on the outflow of traffic along the efferent branches of the autonomic nervous system; one of the Laceys' great insights was to emphasise how central dispatches to the heart are rebounded back to the brain via an abundance of afferents. In stark contrast, most recent cognitive neuroscientific interest in neuro-visceral interaction seems to run almost entirely in the opposite direction (cf. the cycle-timing and synchronisation literature discussed in Chapter 4). While there is certainly nothing wrong with pursuing a bottom-up perspective in itself, any thoroughgoing account of embodiment is ultimately going to have to integrate both arms of the cycle within a single, unified scheme. Active inference would seem to provide an ideal framework for precisely this endeavour.

5.0.3 Covert action under active inference

Before proceeding further, it is worth highlighting that the notion of covert action as discussed thus far is rather more general than that recently introduced into the active inference literature. In the context of active inference, covert action is not a physiological concomitant of a cognitive or mental process, but rather a mode of mental activity in itself. Given the psychophysiological focus of the experiments reported in this thesis, a more general understanding of covert activity as any form of internalised action has been retained (although the apparent discrepancy between these narrower and broader conceptions of covert action might be resolved by construing adaptive physiological

activity as an *expression* of covert action). In any case, it's worth briefly considering the particularities of covert action as understood under active inference.

Mental action was briefly mentioned in Chapter 3 (footnote 45) in relation to counterfactual active inference and covert analogues of epistemic action. This idea related to Pezzulo's (2017) conception of mental action as a form of internal exploration, whereby mental states are deliberately sampled in order to procure new information (or reappraise old information) and thus reduce uncertainty over one's beliefs. Interestingly, this characterisation of mental action, in which one essentially drives one's own train of thought towards a predetermined (epistemic) destination, brings active inference into contact with notions of cognitive control and self-monitoring (cf. Deane 2020; Metzinger 2017; Pezzulo 2012; Pezzulo and Castelfranchi 2009) – a topic further discussed in Chapter 6.

Limanowski and Friston (2018) have since provided a more general characterisation of mental action as the assignment of *expected precision* over subordinate beliefs (i.e. top-down precision-optimisation). This formulation thus conceives of mental action as fundamentally attentional in nature (cf. Feldman and Friston 2010). Building on this idea, Parr, myself, and colleagues (2019) recently proposed an active inference model of the spontaneous fluctuations in visual experience that emerge under conditions of inter-ocular conflict (i.e. binocular rivalry; Levelt 1965; Wheatstone 1838). This model conceptualises the phenomenology of binocular rivalry – an important phenomenon in vision science and consciousness studies – as the consequence of involuntary attentional switches that resolve the uncertainty accumulating in the previously-suppressed (i.e. unattended) visual stream.

While it is not necessary to describe the details of this model in detail here, the key point for the purposes of this chapter is that binocular rivalry can be explained in terms of a fairly simple architecture that (1) embodies the normative principles of active inference, (2) encodes certain beliefs about the environment, and (3) is only able to resolve uncertainty over one sensory channel at a time. The finding that simulations under this scheme engender perceptual fluctuations reminiscent of binocular rivalry implies that cognitive systems that conform to the free energy principle are sensitive to the uncertainty (i.e. perceptual ambiguity) induced by inter-ocular conflict, and what's more, should engage in attentionally-mediated actions in an effort to alleviate it. The aim of the following experiment was to assess whether human subjects evince psychophysiological evidence of analogous covert actions in precisely this scenario.

5.0.4 The present manuscript

The experiment reported below adopts a binocular rivalry paradigm in which incompatible images (faces and houses) were simultaneously presented one to each eye for periods of 1 min. During this time, participants reported the content of their visual experience (i.e. whether they were seeing a face or a house at any given moment) via key press. Each of these ‘rivalry’ trials was followed by a yoked ‘replay’ trial, in which the sequence of percepts reported during the preceding rivalry trial was played back to *both* eyes (i.e. without ocular conflict). Participants were not informed of the difference between trial conditions, and were required to provide the same behavioural responses (i.e. online report of visual experience) in all trials. Cardiac (pulse wave) and electrodermal activity were recorded throughout the experiment and normalised according to a resting-state baseline.

Given the reliable association between states of attentive observation (such as those engendered by the rivalry-replay paradigm) and the modulation of cardiac dynamics (discussed in the Introduction of the manuscript), my co-authors and I set out to test whether the latter constitutes a stereotypic concomitant of sustained attention, or whether cardiac modulation manifests sensitivity to different sources/quantities of uncertainty. The beauty of this experimental setup is that it enabled us to carefully manipulate the kind of uncertainty to which the participant was exposed across different conditions. All trials induced some degree of temporal (or response) uncertainty, by virtue of the unpredictable timing of perceptual transitions. Crucially, however, the inter-ocular conflict induced during rivalry trials engendered additional perceptual ambiguity, while holding the content of perception (and thus, task demands) invariant. Our results indicate that cardiac dynamics are indeed sensitive to these different varieties of uncertainty, lending credence to the notion that the heart may be enlisted as part of a covert, allostatic response to unresolved uncertainty.



Be still my heart: Cardiac regulation as a mode of uncertainty reduction

Andrew W. Corcoran¹ · Vaughan G. Macefield^{2,3} · Jakob Hohwy¹

Accepted: 23 January 2021 / Published online: 23 March 2021
© The Psychonomic Society, Inc. 2021

Abstract

Decreased heart rate (HR) and variability (HRV) are well-established correlates of attention; however, the functional significance of these dynamics remains unclear. Here, we investigate whether attention-related cardiac modulation is sensitive to different varieties of uncertainty. Thirty-nine adults performed a binocular rivalry-replay task in which changes in visual perception were driven either internally (in response to constant, conflicting stimuli; rivalry) or externally (in response to physically alternating stimuli; replay). Tonic HR and high-frequency HRV linearly decreased as participants progressed from resting-state baseline (minimal visual uncertainty) through replay (temporal uncertainty) to rivalry (temporal uncertainty and ambiguity). Time-resolved frequency estimates revealed that cardiac deceleration was sustained throughout the trial period and modulated by ambiguity, novelty, and switch rate. These findings suggest cardiac regulation during active attention may play an instrumental role in uncertainty reduction.

Keywords Heart rate · Attention · Active inference · Multistable perception

Introduction

A growing body of research suggests that physiological signals exert a pervasive influence over cognitive processing (Azzalini et al., 2019; Critchley & Garfinkel, 2018; Quadri et al., 2018). Much of this recent work draws on the predictive coding/active inference formalism (Friston et al., 2017) to explain how the brain integrates interoceptive and exteroceptive information to optimize beliefs about the world. One implication of this perspective is that the

brain not only updates its representations in accordance with accumulating sensory evidence but also regulates its outflows in order to reduce uncertainty (Parr, 2020). The present study pursues this idea in the context of cardiac regulation; specifically, the modulation of heartbeat dynamics classically associated with attentive observation.

Cardiac deceleration (i.e., decreased heart rate; HR) and stabilization (i.e., decreased heart rate variability; HRV) are well-established psychophysiological correlates of attention (Jennings, 1986; Porges, 1992). According to one early, influential view, cardiac deceleration (‘attentional bradycardia’) plays an instrumental role in aiding “both the organism’s receptivity to afferent stimulation and the organism’s readiness to make effective responses to such stimulation” (Lacey, 1972, p. 183). This functional conception of cardiac regulation was proposed on the basis of neurophysiological evidence that arterial baroreceptor stimulation caused by the expulsion of blood from the heart inhibits cortical activity (Lacey & Lacey 1958, 1970). If each influx of baroreceptor feedback does indeed degrade the brain’s capacity to register and respond to external events, reducing the incidence of such perturbations (e.g., by slowing the heartbeat) should facilitate sensorimotor processing (Lacey, 1959, 1967; Lacey & Lacey, 1974).

Despite early criticism of the Laceys’ theoretical claims (e.g., Carroll & Anastasiades, 1978; Elliott, 1972; Hahn,

✉ Andrew W. Corcoran
andrew.corcoran1@monash.edu

Vaughan G. Macefield
vaughan.macefield@baker.edu.au

Jakob Hohwy
jakob.hohwy@monash.edu

¹ Cognition and Philosophy Laboratory, Monash University,
Room E672, 20 Chancellors Walk,
Clayton, VIC 3800, Australia

² Baker Heart and Diabetes Institute, Level 4, 99 Commercial
Road, Melbourne, VIC 3004, Australia

³ Department of Physiology, University of Melbourne,
Parkville, VIC 3010, Australia

1973), a substantial amount of evidence favoring the baroreceptor hypothesis has since accrued. From a neuroanatomical perspective, baroreceptor afferents traveling in the glossopharyngeal and vagus nerves have been shown to project to the nucleus tractus solitarius in the medulla, and thence onto the insular cortex (a central hub for a variety of neural functions, including multimodal integration, attentional orienting, and autonomic control; Gogolla, 2017; Uddin et al., 2017) via the parabrachial nucleus in the pons (Saper & Loewy, 1980). Barosensitive neurones have been found in the insular cortex in the rat and monkey (Zhang & Oppenheimer, 1997; Zhang et al., 1998), and functional magnetic resonance imaging (fMRI) studies in humans have confirmed that baroreceptors project to the insular cortex (Shoemaker et al., 2012).

Although controlled manipulations of baroreceptor activation remain methodologically challenging, several lines of research suggest the central inhibitory effects observed in early animal preparations (Bonvallet et al., 1954; Bonvallet & Allen, 1963; Koch, 1932; Nakao et al., 1956) are likewise operative in intact humans (e.g., Dworkin et al., 1994; Rau et al., 1993; see Elbert & Rau, 1995; Rau & Elbert, 2001, for review). Baroreceptor feedback is now routinely invoked as the key mechanistic explanation of cardiac influences on cognition and behavior (Azzalini et al., 2019; Critchley & Harrison, 2013; Critchley & Garfinkel, 2018; Park & Tallon-Baudry, 2014).

In addition to the bottom-up influence of baroreceptor feedback on brain function, top-down mechanisms of cardiovascular regulation have also been elucidated. After replicating Lacey's (1967, see also Obrist, 1963) observation that attentional bradycardia is accompanied by HRV suppression (Porges & Raskin, 1969), Porges embarked on a programme of research investigating the utility of HRV as an index of parasympathetic (vagal) outflow (see Porges, 1992, 2007, for discussion). This work culminated in Porges' (1995; 2007) seminal polyvagal theory, which subsequently inspired Thayer and Lane's (2000, 2009) neurovisceral integration model. These frameworks organize large bodies of empirical literature around the axis of brain–heart communication, with HRV measures serving to illuminate the tight linkage between physiological, psychological, and behavioral (mal)adaptation. However, despite their foundational interest in the regulation of attentional and autonomic states, these schemes do not explain the role of HRV suppression in sustained attention.

In its latest iteration, the neurovisceral integration model is specified as a recursive, multi-layered control architecture under the active inference formalism (Smith et al., 2017). Active inference provides a computational account of neural function that casts hierarchical brain activity in terms of uncertainty reduction, whereby sensory inputs garnered across multiple modalities are modeled and acted upon

in accordance with the normative principles of Bayesian inference (Friston, 2010; Friston et al., 2017). Adding to recent developments of this framework in the interoceptive domain (e.g., Barrett, 2017; Gu et al., 2019; Khalsa et al., 2018; Owens et al., 2018; Petzschner et al., 2017; Seth & Tsakiris, 2018); Allen and colleagues (2019) have proposed an active inference model of cardio-visual integration that incorporates the baroreceptor hypothesis as a special case. Of particular relevance to our interests here, this approach links perceptual experience and cardiac activity via attentional mechanisms of neuromodulatory gain control (Fardo et al., 2017; Feldman & Friston, 2010). One advantage of this perspective is its potential capacity to reconcile a number of seemingly paradoxical findings within the cardiac cycle-timing literature, in which baroreceptor activation seems to enhance (rather than inhibit) perceptual acuity (see e.g., Garfinkel & Critchley, 2016).

Allen and colleagues' (2019) cardiac active inference model associates the phasic inhibitory effects classically ascribed to baroreceptor activation with transient periods of sensory attenuation (cf. Brown et al., 2013; Limanowski, 2017). This implies that perceptual sensitivity depends on the prevailing heart rhythm, which dictates the number of heartbeat-related attenuation events that occur over a given timeframe. This perspective thus supports the notion that cardiac activity can be instrumentally regulated to optimize perceptual processing, where cardiac deceleration reduces the frequency of attenuation events and thereby increases the opportunity for high-precision sensory sampling. Moreover, this account might also clarify why attention to external events is typically accompanied by HRV suppression: if distributing heartbeats more regularly over time enables the brain to predict the timing of baroreceptor impulses with greater precision (cf. Al et al., 2020; Porges, 1992), this may enhance its ability to minimize their impact on exteroceptive processing.

If cardiac regulation does conform to the principles of active inference, this implies that attention-related heartbeat activity derives from a fundamental imperative to resolve uncertainty. Accordingly, this study set out to investigate whether cardiac dynamics are sensitive to different varieties of uncertainty. To this end, we contrasted autonomic measures recorded during a binocular rivalry–replay task designed to elicit unpredictable oscillations in visual awareness. Since this task involved sustained periods of focused attention, we expected it to evoke cardiac deceleration and stabilization relative to the resting state. Crucially, we expected these effects to be stronger during rivalry trials, which involved an additional source of uncertainty (i.e., ambiguity) over and above the temporal (response) uncertainty common to both conditions. As replay mimicked the phenomenology and behavioral

demands of rivalry, this finding would provide compelling evidence that cardiac activity is modulated by uncertainty rather than the deployment or capture of attention *per se*.

Materials and methods

Participants

Contemporary psychophysiological studies of cardio-sensory integration typically involve sample sizes on the order of 30–40 participants (e.g., Al et al., 2020; Azevedo et al., 2018; Galvez-Pol et al., 2020; Hodossy & Tsakiris, 2020; Makowski et al., 2020; Motyka et al., 2019). Accordingly, 39 adults (26 female) aged 18–46 years ($M = 22.46$ years, $S.D. = 4.59$) were recruited via a university-wide research participant pool.

All participants were fluent English speakers who reported normal (or corrected-to-normal) vision, no history of chronic illness, neurological, psychological, cardiovascular, respiratory, metabolic, endocrine, immune, or substance abuse disorder, no concussion or unexplained loss of consciousness > 1 min, and no intellectual impairment or learning disability. Participants reported no tobacco or illicit drug use within the previous 6 months, no regular medication use (contraception excepted), and were not pregnant, lactating, or engaged in elite-level physical training at the time of the experiment. Four participants were left-handed; 11 were left-eye dominant (hole-in-the-card test; Dolman, 1919).

Participants provided written, informed consent and were remunerated AU\$20 for their time. This study was

approved by the Monash University Human Research Ethics Committee (ID: 13966).

Psychophysical stimuli and apparatus

Visual stimuli were generated with the Psychophysics Toolbox (v3; Brainard, 1997) in MATLAB R2013a (v8.1.0.604; The MathWorks Inc., Natick, MA, USA). Participants viewed stimuli via a mirror stereoscope (ScreenScope SA200LT, StereoAids, Albany, WA) while sitting with their head stabilized on a chinrest 43 cm from the screen. Perceptual reports were recorded via a gaming controller.

Calibration stimuli were black and white checkerboards subtending $\sim 5.1^\circ$ of visual angle presented on a uniform grey background (Fig. 1a). These stimuli also served as vergence cues during the task, along with central fixation crosses subtending $\sim 0.4^\circ$.

Task stimuli were eight face and eight house photographs sampled from the Chicago Face Database (Ma et al., 2015) and DalHouses stimulus set (Filliter et al., 2016), respectively (see [Supplementary Materials](#) for details). Images were sorted into face–house pairs and matched for luminance and spatial frequency using the SHINE toolbox (Willenbockel et al., 2010). They were presented through an oval aperture subtending $\sim 4.1^\circ \times 2.9^\circ$.

Procedure

Sessions began with eligibility screening, consent procedures, and a short demographic survey. A brief calibration

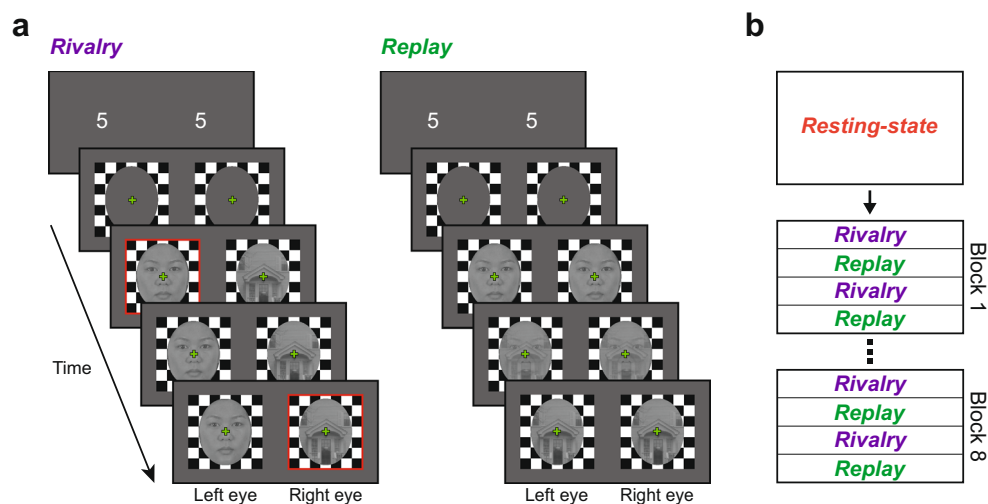


Fig. 1 Experimental design. **a.** Schematic of rivalry-replay trial pair. Each trial was preceded by a 5-s countdown and a 1-s vergence cue. In rivalry trials (left), face and house stimuli were continuously presented to one eye each for a period of 60 s. Replay trials (right) presented identical stimuli to both eyes in accordance with the time series of dominant percepts reported on the preceding rivalry trial (outlined here in red); overlaid stimuli were presented in phases between dominant percepts. **b.** Schematic of experimental condition/trial order

procedure was then undertaken in which identical checkerboard stimuli were presented to both eyes. The horizontal distance between these stimuli was adjusted until the participant reported convergence. Task stimuli were subsequently presented at the center of these checkerboards.

Next, participants performed one or two practice blocks of the experimental task (described below) as required. This was followed by a 5-min resting-state recording, in which participants were instructed to sit quietly and relax with their eyes closed. The experimenter vacated the room during this recording.

The experimental task consisted of eight blocks of interleaved rivalry and replay trials (Fig. 1b). Participants were not informed about the difference between these conditions. Each block comprised four 60-s trials featuring one stimulus pair. Trials were cued with a 5-s onscreen countdown, followed by a 1-s checkerboard and fixation cross. These vergence cues remained present for the remainder of the trial. Block order was randomized across participants. Self-paced breaks were available between blocks.

During rivalry, paired stimuli were presented to one eye each, inducing a visual stream that spontaneously oscillated between face and house percepts. Stimulus laterality was randomized on the first trial and reversed on the second rivalry trial. Participants were instructed to fixate on the cross and press one of two response keys for as long as they perceived $\geq 90\%$ of either stimulus (button-stimulus mapping counterbalanced across participants). Periods in which no responses were registered were classed as transition (i.e., mixed or ‘piecemeal’) phases.

In replay trials, identical sequences of alternating stimuli were presented to both eyes. Sequences were generated from the participant’s subjective report on the preceding trial (Lumer et al., 1998; onset/offset latencies shifted -400 ms to account for response latency; Knapen et al., 2011). Transitions between dominant phases were realized via a ‘cross-fading’ technique, whereby the transparency of superposed images was linearly increased/decreased over the course of 400 ms. For longer transition periods, stimuli were suspended at 50% transparency until the final 200 ms of the interval (Gelbard-Sagiv et al., 2018). Phases < 400 -ms duration interposed between identical percepts were interpolated to render a single continuous percept during the corresponding replay period.

Electrophysiological signal acquisition and preprocessing

Pulse photoplethysmograph (PPG) and skin potential records were acquired in LabChart (v7.3.8; ADInstruments, Bella Vista, NSW) using a PowerLab 26T system. PPGs were recorded bilaterally from the second toes; skin

potentials were recorded from Ag/AgCl electrodes on the dorsal and plantar surfaces of each foot (left medial malleolus as reference). Signals were digitized at 1000 Hz with online high-pass (0.5 Hz, -3 dB cutoff) and anti-aliasing hardware filters applied. Pulse amplitude was estimated from each PPG during the online recording.

Resting-state and task records were exported to MATLAB R2019b (v9.7.0.1319299) for offline processing. Records were visually inspected and artifact-contaminated signals rejected. Data were epoched into trial periods using the EEGLAB toolbox (v2019.1; Delorme & Makeig, 2004). Pulse amplitude and skin potential estimates were z-normalized to the resting-state recording. Estimates $\pm 5S.D.$ from the sample mean were excluded.

Mean inter-beat interval (IBI; reciprocal of HR) and high-frequency HRV (HF-HRV; a.k.a. respiratory sinus arrhythmia; RSA) were estimated from systolic pressure wave peaks using semi-automated classification software (ARTiiFACT v2.09; Kaufmann et al., 2011). IBI time-series for non-overlapping 60-s segments of resting-state records and trial epochs were extracted from the PPG with the least missing data (left channel if tied). Records were visually inspected to ensure systolic peaks had been accurately resolved. Artifactual IBIs were corrected with cubic-spline interpolation prior to fast Fourier transform (default parameters).

Time-resolved estimates of instantaneous heart frequency (IHF) were extracted from wavelet-transformed PPGs (frequency range = $0.5 - 2 \times$ mean heart frequency) using an adaptive ridge curve technique (NMD Toolbox, v2.00; Iatsenko et al., 2015; Iatsenko et al., 2016). This method uses the entire PPG time series to reconstruct IHF (rather than relying solely on the information provided by discrete peak events; cf. Iatsenko et al., 2013), thus rendering a near-continuous, high-resolution representation of time-evolving cardiac dynamics. Trials containing < 15 s of data were excluded from analysis (four trials from two participants each). IHF estimates were z-normalized to the resting-state recording.

Data analysis

All statistical analyses were conducted in R (v3.6.2; R Core Team, 2019) with the RStudio Desktop IDE (v1.2.5033; RStudio Team, 2015). We used linear mixed-effects models (LMMs) to estimate the mean IBI and (natural log-transformed) HF-HRV expected across conditions. IBI was modeled in favor of HR because (1) fluctuations in IBI are more linearly related to parasympathetic outflow than HR (Berntson et al., 1995; Quigley & Berntson, 1996), and (2) IBIs form the basic unit of measurement from which HRV estimates are derived. Back-transformed estimates of mean HR (in beats per minute; bpm) are provided

alongside IBI results in order to aid interpretability. Since IBI and HF-HRV are typically correlated (de Geus et al., 2019), they were included as covariates in one another's models. Condition was entered as an ordered factor (*rest* > *replay* > *rivalry*). By-participant random intercepts were included to account for repeated measures.

We additionally modeled switch rate as a function of task condition, mean IBI, and HF-HRV. Switch rate was calculated as the number of dominant percepts per trial/trial duration (s). This model included two-way interactions between condition and IBI, and condition and HF-HRV. Random intercepts were included for participant and trial-pair identity, where the latter accounted for variation in the time-course of perceptual alternations deriving from the interaction between stimulus identity and eye dominance.

Next, we used generalized additive mixed-effects models (GAMMs) to estimate the expected time-course of IHF, pulse amplitude, and skin potential fluctuations over the trial period (downsampled to 2 Hz). These models enabled us to assess the functional form of electrophysiological responses to differing experimental conditions. Since these models capture information about the mean and variance of physiological changes over time, response variables were z-normalized at the subject-level to account for individual differences in resting-state activity and lability. Negative-valued model predictions thus reflect a decrease in physiological activity (e.g., lower IHF) during task performance relative to resting-state baseline. A 1-unit deviation indicates that the magnitude of this change was equivalent to 1 standard deviation from the resting-state mean. Note that identically specified models fit to mean-centered data revealed qualitatively similar results, albeit with lower fit indices (see [Supplementary Materials](#)). To aid interpretability, estimates of HR change based on these models are provided alongside normalized estimates.

LMMs and GAMMs were estimated using the *mgcv* package (v1.8-31; Wood, 2011). Smooth functions for fixed effects were fit using low-rank thin-plate regression splines (Wood, 2003, 2017). Factor smooths (i.e., nonlinear random effects) were specified for both participant and trial-pair identity over trial time (see Baayen et al., 2017; Cross et al., 2020, for similar approaches). These terms were fit with a first-derivative penalty to shrink them towards the population-level. Models were fit with an AR(1) process using the *itsadug* package (v2.3; van Rij et al., 2017), and assessed using the *mgcViz* package (v0.1.6; Fasiolo et al., 2018).

All reported parametric effects are estimated marginal means obtained (and statistically evaluated) via the *emmeans* package (v1.4.5; Lenth, 2020). Model visualization was aided by the *tidyverse* (v1.3.0; Wickham et al., 2019) and *ggpubr* (v0.2.5; Kassambara, 2020) packages.

Results

Behavioral performance

One block of data was missing due to incorrect button presses. Mean duration of median dominant percepts was 4.96 s ($SD = 4.39$). Mean predominance score ($\sum Duration_{face} / \sum Duration_{face} + Duration_{house}$) was 0.70 ($SD = 0.13$), indicating a bias towards faces. Dominant percepts alternated at a mean switch rate of 0.16 Hz ($SD = 0.08$). These measures were highly correlated with their replay condition counterparts ($\rho = 0.88, 0.88$, and 0.96 , respectively), suggesting replay evoked similar response profiles to those engendered in rivalry.

Inter-beat interval and heart rate variability

LMMs revealed significant linear contrasts across conditions for mean IBI and HF-HRV (Table 1). Mean IBI increased by an average 12.56 ms (-1.32 bpm) during replay, $t(1399) = 4.21, p < .001$, and 20.45 ms (-2.12 bpm) during rivalry, $t(1399) = 6.85, p < .001$, relative to resting-state levels. Conversely, HF-HRV decreased by an average -0.09 units during replay, $t(1400) = 1.86, p = .150$, and -0.11 units during rivalry, $t(1400) = 2.44, p = .040$, compared to resting-state (Fig. 2). In line with our predictions, these findings suggest that heightened perceptual uncertainty evokes slower, more regular heartbeats.

Switch rates did not significantly differ as a function of condition, IBI, HF-HRV, or interactions amongst these variables (see [Supplementary Materials](#) for details). The lack of association between switch rate and mean IBI accords with previous reports (Hodges & Fox, 1965; Shannon et al., 2011).

Instantaneous heart frequency

In line with the IBI model, time-resolved IHF estimates were significantly lower during rivalry than replay (Table 2). Visualization of model smoothers revealed that IHF tended to decline steeply over the first 10 s of the trial period, nadired around 11–12 s, and undulated below baseline for the remainder of the trial. IHF declined by -0.90 n.u. (~ 3.94 bpm) during rivalry (compared to -0.58 n.u. [~ 2.39 bpm] during replay), before recovering to a level commensurate with that of replay. IHF was significantly lower in rivalry for all but the final 10 s of the trial period (Fig. 3).

Since rivalry–replay paradigms necessitate a fixed presentation order in which rivalry always precedes replay, we analyzed whether differences between conditions derived from an order effect (Table 2). Refitting the GAMM with time-by-trial smoothers revealed that cardiac

Table 1 Summary of linear mixed-effects models for inter-beat interval and high-frequency heart rate variability

Inter-beat interval					Heart rate variability				
Predictors	Estimate	S.E.	t	p	Predictors	Estimate	S.E.	t	p
(Intercept)	763.32	14.91	51.18	< .001	(Intercept)	5.77	0.13	44.75	< .001
Condition.L	14.46	2.11	6.85	< .001	Condition.L	-0.08	0.03	-2.44	.015
Condition.Q	-1.91	1.38	-1.39	.165	Condition.Q	0.02	0.02	1.01	.312
scale(HRV)	18.75	1.48	12.67	< .001	scale(IBM)	0.62	0.04	14.06	< .001
Condition.L: scale(HRV)	0.16	1.74	0.09	.927	Condition.L: scale(IBM)	-0.04	0.03	-1.18	.238
Condition.Q: scale(HRV)	2.12	1.21	1.76	.079	Condition.Q: scale(IBM)	0.03	0.02	1.36	.174
<i>Smoothers</i>	<i>e.d.f.</i>	<i>d.f.</i>	<i>F</i>	<i>p</i>	<i>Smoothers</i>	<i>e.d.f.</i>	<i>d.f.</i>	<i>F</i>	<i>p</i>
s(ID _{subj})	37.83	38	198.30	< .001	s(ID _{subj})	37.39	38	73.53	< .001
Adjusted R ²	.92				Adjusted R ²	.80			
Observations	1443				Observations	1443			

Cardiac parameters are included as covariates to account for shared variance and are thus not interpreted. L = linear contrast; Q = quadratic contrast. *e.d.f.* = estimated degrees of freedom; associated *p* values are only approximate. *ID_{subj}* smoothers are equivalent to by-participant random intercepts.

deceleration was most pronounced on the first trial (Fig. 4). Since each block comprised a unique stimulus set, we ascribe this effect to stimulus novelty. Crucially, the second rivalry trial was associated with greater IHF reduction than the *preceding* replay trial for the majority of the trial period. This result argues against a simple order effect whereby cardiac deceleration diminished over the course of the block, or where condition-level differences were driven by the unique response profile of the first trial.

Finally, we examined whether time-evolving IHF differed between conditions as a function of switch rate. Including switch rate as a tensor product smoother over time revealed that the by-condition difference in IHF declined as the number of perceptual alternations increased (Fig. 5). More specifically, as switch rate increased, IHF in rivalry ceased to be significantly lower than in replay at a progressively earlier time-point. We speculate that the deeper deceleration elicited by rivalry is sustained for longer when the ambiguity induced by inter-ocular conflict requires more time to resolve (see Hohwy et al., 2008; Parr et al., 2019).

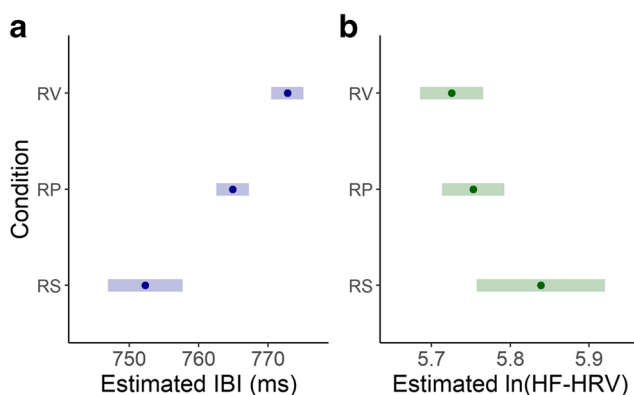


Fig. 2 Linear mixed-effects model estimates for cardiac parameters. Points indicate trial-level estimated marginal means for (a) mean inter-beat interval (IBI) and (b) log-transformed high-frequency heart rate variability (HF-HRV) across resting-state (RS), replay (RP), and rivalry (RV) conditions. Shading indicates 95% confidence intervals

Pulse wave and skin potential amplitude

Average pulse amplitude showed no significant deviation from resting-state levels, $t(291681) = 1.10$, $p = .273$, and no significant difference between conditions, $t(291681) = 1.22$, $p = .222$. Time-by-condition smoothers were non-significant (see [Supplementary Materials](#) for details). Hence, there was no evidence of any systematic difference in vasomotor response across conditions.

Average skin potential amplitude was significantly decreased in rivalry relative to replay, $t(291974) = 3.01$, $p = .003$. Decomposition by trial order revealed that this difference was driven by a pronounced reduction in skin potential amplitude during the first trial; the magnitude of this effect diminished over successive trials (see [Supplementary Materials](#) for details). Differences in

Table 2 Summary of generalized additive mixed-effects models for instantaneous heart frequency

IHF by Condition					IHF by Trial				
Predictors	Estimate	S.E.	t	p	Predictors	Estimate	S.E.	t	p
(Intercept)	-0.58	0.19	-3.09	.002	(Intercept)	-0.58	0.19	-3.13	.002
Replay	0.10	0.01	10.60	< .001	Rivalry1	-0.21	0.02	-12.95	< .001
					Replay1	0.06	0.02	3.76	< .001
					Rivalry2	0.01	0.02	0.70	0.482
Smoothers	e.d.f.	d.f.	F	p	Smoothers	e.d.f.	d.f.	F	p
s(Time):Rivalry	8.24	8.51	29.56	< .001	s(Time):Rivalry1	8.09	8.56	20.82	< .001
s(Time):Replay	8.40	8.62	40.74	< .001	s(Time):Replay1	8.04	8.53	19.34	< .001
s(Time, ID _{subj})	286.83	350	46.90	< .001	s(Time):Rivalry2	8.52	8.80	36.70	< .001
s(Time, ID _{stim})	111.59	143	3.87	< .001	s(Time):Replay2	7.68	8.29	21.19	< .001
					s(Time, ID _{subj})	286.31	350	47.48	< .001
					s(Time, ID _{stim})	110.74	143	3.79	< .001
Adjusted R ²	.54				Adjusted R ²	.55			
Observations	293729				Observations	293729			

Predictors are sum-to-zero contrast-coded; i.e., estimates reflect deviation from the grand mean (intercept). *e.d.f.* = estimated degrees of freedom; associated *p* values are only approximate

sudomotor activity across conditions may therefore reflect an arousal response to stimulus novelty.

Discussion

Renewed interest in brain–heart communication has generated a wealth of data on the impact of cardio-afferent signaling on conscious awareness and action. Inspired by

recent work formalizing the role of uncertainty in the relation between neural and visceral states (e.g., Allen et al., 2019; Petzschner et al., 2017; Pezzulo et al., 2015; Smith et al., 2017; Stephan et al., 2016), we investigated how cardiac parameters vary as a function of perceptual uncertainty. As predicted, mean inter-beat interval linearly increased, and evinced less variability, from a state of minimal visual uncertainty to one involving unpredictability (stochastic perceptual transition) and ambiguity (inter-ocular conflict). Modeling the time-course of cardiac dynamics confirmed that decreases in tonic heart rate emanated from a sustained deceleration over the course of the trial period. The absence of comparable patterning in two indices of sympathetic out-flow suggests these effects are unlikely to stem from generic fluctuations in arousal.

Further analysis revealed that cardiac deceleration interacted with trial order, an effect we ascribe to the presentation of novel stimuli at the beginning of each block. Cardiac deceleration is a well-known component of the orienting reflex (Graham & Clifton, 1966; Graham, 1979), a constellation of neurophysiological, autonomic, and behavioral adaptations evoked by novel or changing stimuli (Berlyne, 1960; Sokolov, 1960, 1969). Analogous to the way orienting responses habituate to repeated stimuli, we interpret the reduction in deceleration from the first to the second rivalry trial as reflecting a reduction of uncertainty over stimulus features. More concretely, we posit that participants began each block with prototypical face and house models that were subsequently elaborated (i.e., ‘filled-in’) as a consequence of perceptual learning over the

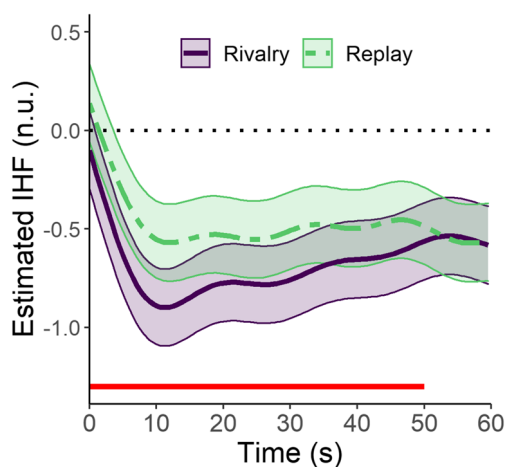


Fig. 3 Time-resolved estimates of instantaneous heart frequency across conditions. The *dotted line* indicates mean instantaneous heart frequency (IHF) during resting-state; negative IHF estimates represent cardiac deceleration relative to this baseline. The *solid red line* indicates significant difference between smoothers. *Shading* indicates standard error of the mean. *n.u.* = normalized units

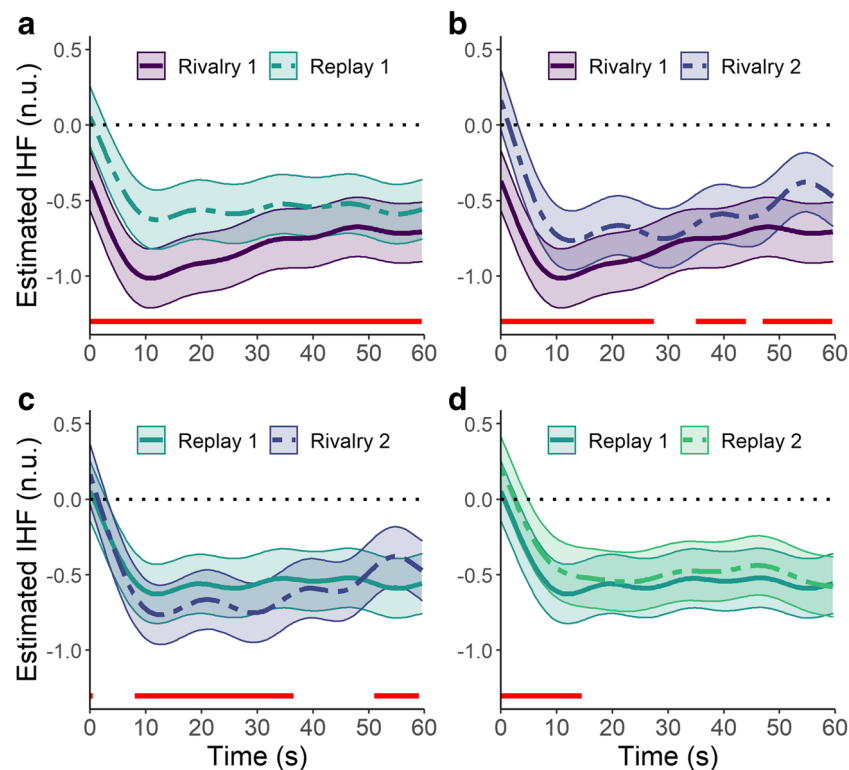


Fig. 4 Pairwise comparisons of instantaneous heart frequency across trials. The *dotted lines* indicate mean instantaneous heart frequency (IHF) during resting-state; negative IHF estimates represent cardiac deceleration relative to this baseline. The *broken lines* depict the later trial within each pair. *Solid red lines* indicate significant differences between smooths. *Shading* indicates standard error of the mean. The first rivalry trial evokes significantly lower IHF than both the first replay trial (**a**) and the second rivalry trial (**b**). Notably, the second rivalry trial elicits lower IHF than the first replay trial for the majority of the trial period (**c**). The second replay trial demonstrates the shallowest decline in IHF over the early phase of the trial period, but is otherwise similar to the response on the first replay trial (**d**). *n.u.* = normalized units

course of the first trial pair. Having reduced uncertainty about the underlying configuration of competing sensory inputs, participants could then harness these models to disambiguate percepts during the second rivalry trial.

The difference between IHF curves on each condition was also found to be modulated by perceptual switch rate. The deeper deceleration response evoked by rivalry returned to a level comparable with that of replay faster (i.e., earlier in the trial) as switch rate increased. This observation indicates that cardiac dynamics are not only sensitive to perceptual uncertainty at a broadly contextual level (i.e., whether or not one is experiencing inter-ocular conflict), but also at the more fine-grained level of perceptual state dynamics (i.e., how rapidly inter-ocular conflict is being resolved). Under the assumption that the ongoing suppression of visual input results in the accumulation of prediction error, lower switch rates imply slower resolution of uncertainty over time; hence, the protracted cardiac deceleration response observed at lower switch rates might reflect an adaptation to accumulating uncertainty. Whether

the relevant source of this uncertainty pertains to the level of perceptual inference, or to higher-order expectations about the rate of prediction error minimization over time, remains to be clarified.

As alluded to in the Introduction, cardiac deceleration and stabilization might indirectly facilitate sensorimotor processing by rendering baroreceptor feedback less frequent and more predictable over time. One way the brain might realize such patterning is by specifying predictions about the timing of each heartbeat that are subsequently enforced by autonomic reflex arcs (Petzschn et al., 2017; Pezzulo et al., 2015; Seth & Friston, 2016; Stephan et al., 2016). Assuming this regimen can be implemented with sufficient temporal precision, each influx of baroreceptor afferent traffic should closely align with its scheduled arrival, thus enabling the brain to attenuate the impact of these signals on exteroceptive processing. While speculative, this account accrues support from evidence that stimulus trains take longer to break into conscious awareness when phase-locked to the heartbeat—an observation that suggests the

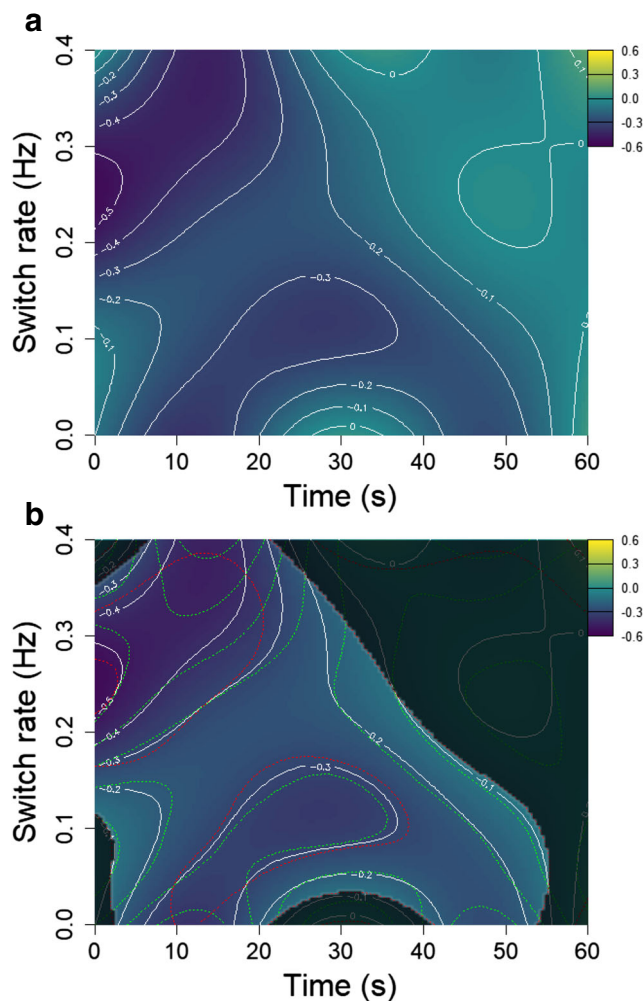


Fig. 5 Modulation of instantaneous heart frequency by switch rate. The *top panel (a)* depicts differences in instantaneous heart frequency between conditions as a function of trial time and perceptual switch rate (green indicates no difference between conditions; darker blue indicates greater deceleration in rivalry). The *bottom panel (b)* masks regions of non-significant difference and includes 95% confidence intervals for contours (red and green dotted lines)

brain may ‘tag’ sensations occurring within the cardiac frequency channel as self-generated (and thus, appropriate to ignore; Salomon et al., 2016).

An intriguing corollary of this arrangement is that it would seem to confer an intrinsic mode of uncertainty reduction *irrespective* of its consequences for exteroceptive processing. By enacting an autonomic (more specifically, *allostatic*; see Corcoran & Hohwy, 2018; Corcoran et al., 2020) policy that renders the sensory consequences of cardiac activity more predictable, the brain thereby helps itself to a precise source of sensory feedback that implicitly resolves uncertainty about the evolving trajectory of internal bodily states (cf. Perrykkad & Hohwy, 2020). It follows that the brain should increasingly exploit such covert modes

of uncertainty reduction when confronted with mounting or persistent uncertainty—a prediction corroborated by our observation that cardiac deceleration was enhanced under novelty and lower switch rates. Recent evidence that HF-HRV suppression is amplified by false cardiac biofeedback (Hodossy & Tsakiris, 2020) is also commensurate with this view, if one interprets the unresolvable discrepancy between incongruent interoceptive and visual signals as uncertainty-accruing.

While the magnitude of the cardiac responses observed during rivalry and replay were relatively small, such effect sizes are in keeping with the broader literature. Classic studies involving vigilance (Bowers, 1971; Lacey et al., 1963; Libby et al., 1973) and visual search tasks (Coles, 1972) report average decreases in tonic HR of <3 bpm. Cardiac orienting studies demonstrate phasic deceleration effects up to ~1 bpm in response to neutral images, and ~2.5–3.5 bpm for complex or unpleasant images (Abercrombie et al., 2008; Bradley et al., 2001; Fredrikson & Öhman, 1979). Moreover, deceleration curves following an auditory cue differ on average by up to ~1 bpm for trials in which a near-threshold visual stimulus is consciously perceived, as opposed to going undetected (Cobos et al., 2019; see also Park et al., 2014). Together, these (and many other) findings imply that the effects of the present study are of a sufficient magnitude to distinguish functionally relevant differences in attentional and perceptual processing.

One limitation of the rivalry–replay paradigm is its lack of ecological validity (for discussion, see Arnold, 2011). Importantly, however, adopting this method enabled us to rigorously match phenomenology and behavior across task conditions, and thus to isolate the effects of perceptual ambiguity from those of sustained attention/vigilance, perceptual alternation, and motor activity. In this way, we were able to demonstrate that cardiac regulation is not only sensitive to different dimensions of uncertainty (i.e., novelty, ambiguity, unpredictability), but also to the *degree* of uncertainty associated with perceptual states (e.g., novel vs. familiar rivalry). These data are consistent with an active inference account of cardio-visual integration in which episodes of baroreceptor feedback are increasingly postponed (within physiological limits) as perceptual inferences accrue greater uncertainty. This implies that the cardiac correlates of attentional orienting, vigilance, and motor preparation emerge from a more fundamental biological imperative to optimize evidence accumulation and resolve uncertainty (Feldman & Friston, 2010; Parr & Friston, 2017). Thus, even though the data reported here were generated under non-ecologically valid conditions, they lend support to the broader hypothesis that perceptual and physiological cycles of inference and adaptation are interwoven through domain-general principles of uncertainty reduction.

Although we interpret our findings as an expression of fundamental regulatory processes, the generalizability of these results beyond healthy, young adults remains an open question. Individual differences may impose significant physiological or computational constraints on the capacity to adapt cardiac states to uncertainty. For instance, older adults show reduced cardiac deceleration during motor preparation (Ribeiro & Castelo-Branco, 2019), an observation that might be explained by a diminished sensitivity to uncertainty (Moran et al., 2014; Nassar et al., 2016). Similarly, blunted HRV might limit the extent to which heartbeat dynamics can be recruited to resolve uncertainty. Indeed, the link between decreased HRV, hyper-vigilance, and affective disorder (Beauchaine & Thayer, 2015; Mulcahy et al., 2019; Park & Thayer, 2014) might suggest that HRV is chronically suppressed as a consequence of unrelenting uncertainty (cf. Ottaviani, 2018; Peters et al., 2017), or alternatively, that a reduced capacity to resolve uncertainty through cardiac regulation may increase vulnerability to psychopathology. While speculative, these remarks highlight the potential utility of viewing the nexus between psychophysiology and psychopathology through the lens of uncertainty reduction.

To conclude, our findings indicate that the cardiac correlates of attention derive from a fundamental biological imperative to resolve uncertainty. These results complement the growing body of research concerning the impact of interoceptive states on consciousness and cognition by highlighting the cyclical, temporally extended nature of brain–heart communication. Future work would benefit from exploring other ways in which the internal milieu may afford opportunities for uncertainty reduction, and how such allostatic dynamics vary in health and disease.

Supplementary Information The online version contains supplementary material available at (<https://doi.org/10.3758/s13423-021-01888-y>).

Acknowledgements This work was supported by the Australian government (RTP scholarship to AWC) and the Australian Research Council (grant numbers DP160102770, DP190101805 to JH). We are grateful to all who participated in this study. We also thank Zachariah Cross, Stephen Gadsby, Kelsey Perrykkad, and two reviewers for feedback that improved the quality of the final manuscript, Bryan Paton and Mateusz Woźniak for invaluable technical advice and assistance, and Phillip Alday for many enlightening discussions on the intricacies of statistical modeling.

Open practices statement This experiment was not formally preregistered. De-identified raw and preprocessed data, along with source code for the analyses presented in this paper, are openly available from the Bridges research repository (Corcoran et al. 2021; <https://doi.org/10.26180/c.5084189>).

References

- Abercrombie, H. C., Chambers, A. S., Greischar, L., & Monticelli, R. M. (2008). Orienting, emotion, and memory: Phasic and tonic variation in heart rate predicts memory for emotional pictures in men. *Neurobiology of Learning & Memory*, 90(4), 644–650.
- Al, E., Iliopoulos, F., Forschack, N., Nierhaus, T., Grund, M., Motyka, P., ..., Villringer, A. (2020). Heart–brain interactions shape somatosensory perception and evoked potentials. *Proceedings of the National Academy of Sciences*, 117(19), 10575–10584.
- Allen, M., Levy, A., Parr, T., & Friston, K. J. (2019). In the body's eye: The computational anatomy of interoceptive inference. *bioRxiv*.
- Arnold, D. H. (2011). Why is binocular rivalry uncommon? Discrepant monocular images in the real world. *Frontiers in Human Neuroscience*, 5, 116.
- Azevedo, R. T., Badoud, D., & Tsakiris, M. (2018). Afferent cardiac signals modulate attentional engagement to low spatial frequency fearful faces. *Cortex*, 104, 232–240.
- Azzalini, D., Rebollo, I., & Tallon-Baudry, C. (2019). Visceral signals shape brain dynamics and cognition. *Trends in Cognitive Sciences*, 23(6), 488–509.
- Baayen, H. R., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory & Language*, 94, 206–234.
- Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive & Affective Neuroscience*, 12(1), 1–23.
- Beauchaine, T. P., & Thayer, J. F. (2015). Heart rate variability as a transdiagnostic biomarker of psychopathology. *International Journal of Psychophysiology*, 98(2 Pt 2), 338–350.
- Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. New York, Toronto, London: McGraw-Hill Book Company.
- Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1995). The metrics of cardiac chronotropism: Biometric perspectives. *Psychophysiology*, 32(2), 162–171.
- Bonvallet, M., & Allen, J. (1963). Prolonged spontaneous and evoked reticular activation following discrete bulbar lesions. *Electroencephalography & Clinical Neurophysiology*, 15, 969–988.
- Bonvallet, M., Dell, P., & Hiebel, G. (1954). Tonus sympathique et activité électrique corticale. *Electroencephalography & Clinical Neurophysiology*, 6, 119–144.
- Bowers, K. S. (1971). Heart rate and GSR concomitants of vigilance and arousal. *Canadian Journal of Psychology*, 25(3), 175–184.
- Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, 1(3), 276–298.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- Brown, H., Adams, R. A., Parees, I., Edwards, M., & Friston, K. J. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing*, 14(4), 411–427.
- Carroll, D., & Anastasiades, P. (1978). The behavioural significance of heart rate: The Lacey's hypothesis. *Biological Psychology*, 7(4), 249–275.
- Cobos, M. I., Guerra, P. M., Vila, J., & Chica, A. B. (2019). Heart-rate modulations reveal attention and consciousness interactions. *Psychophysiology*, 56(3), e13295.
- Coles, M. G. H. (1972). Cardiac and respiratory activity during visual search. *Journal of Experimental Psychology*, 96(2), 371–379.

- Corcoran, A. W., & Hohwy, J. (2018). Allostasis, interoception, and the free energy principle: Feeling our way forward. In Tsakiris, M., & De Preester, H. (Eds.) *The interoceptive mind: From homeostasis to awareness*, Oxford: Oxford University Press, pp 272–292.
- Corcoran, A. W., Pezzulo, G., & Hohwy, J. (2020). From allostatic agents to counterfactual cognisers: Active inference, biological regulation, and the origins of cognition. *Biology & Philosophy*, 35(32) pp. 1–45.
- Corcoran, A. W., Macefield, V. G., & Hohwy, J. (2021). Be still my heart: Cardiac regulation as a mode of uncertainty reduction. Monash University. Collection. <https://doi.org/10.26180/c.5084189>.
- Critchley, H. D., & Garfinkel, S. N. (2018). The influence of physiological signals on cognition. *Current Opinion in Behavioral Sciences*, 19, 13–18.
- Critchley, H. D., & Harrison, N. A. (2013). Visceral influences on brain and behavior. *Neuron*, 77(4), 624–638.
- Cross, Z. R., Corcoran, A. W., Schlesewsky, M., Kohler, M. J., & Bornkessel-Schlesewsky, I. (2020). Oscillatory and aperiodic neural activity jointly predict grammar learning. *bioRxiv*.
- de Geus, E. J. C., Gianaros, P. J., Brindle, R. C., Jennings, J. R., & Bertson, G. G. (2019). Should heart rate variability be “corrected” for heart rate? Biological, quantitative, and interpretive considerations. *Psychophysiology*, 56(2), e13287.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open-source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21.
- Dolman, P. (1919). Tests for determining the sighting eye. *American Journal of Ophthalmology*, 2, 867.
- Dworkin, B. R., Elbert, T., Rau, H., Birbaumer, N., Pauli, P., Droste, C., & Brunia, C. H. (1994). Central effects of baroreceptor activation in humans: Attenuation of skeletal reflexes and pain perception. *Proceedings of the National Academy of Sciences*, 91(14), 6329–6333.
- Elbert, T., & Rau, H. (1995). What goes up (from heart to brain) must calm down (from brain to heart)! Studies on the intersection between baroreceptor activity and cortical excitability. In Vaitl, D., & Schandry, R. (Eds.) *From the heart to the brain: The psychophysiology of circulation-brain interaction*, Frankfurt am Main: Peter Lang, pp 133–149.
- Elliott, R. (1972). The significance of heart rate for behavior: A critique of Lacey’s hypothesis. *Journal of Personality & Social Psychology*, 22(3), 398–409.
- Fardo, F., Aukstulewicz, R., Allen, M., Dietz, M. J., Roepstorff, A., & Friston, K. J. (2017). Expectation violation and attention to pain jointly modulate neural gain in somatosensory cortex. *NeuroImage*, 153, 109–121.
- Fasiolo, M., Nedellec, R., Goude, Y., & Wood, S. N. (2018). Scalable visualisation methods for modern generalized additive models. *arXiv:1809.10632*.
- Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4(215), 1–23.
- Filliter, J. H., Glover, J. M., McMullen, P. A., Salmon, J. P., & Johnson, S. A. (2016). The DalHouses: 100 new photographs of houses with ratings of typicality, familiarity, and degree of similarity to faces. *Behavior Research Methods*, 48(1), 178–183.
- Fredrikson, M., & Öhman, A. (1979). Heart-rate and electrodermal orienting responses to visual stimuli differing in complexity. *Scandinavian Journal of Psychology*, 20(1), 37–41.
- Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.
- Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation*, 29(1), 1–49.
- Galvez-Pol, A., McConnell, R., & Kilner, J. M. (2020). Active sampling in visual search is coupled to the cardiac cycle. *Cognition*, 196, 104149.
- Garfinkel, S. N., & Critchley, H. D. (2016). Threat and the body: How the heart supports fear processing. *Trends in Cognitive Sciences*, 20(1), 34–46.
- Gelbard-Sagiv, H., Mudrik, L., Hill, M. R., Koch, C., & Fried, I. (2018). Human single neuron activity precedes emergence of conscious perception. *Nature Communications*, 9(1), 2057.
- Gogolla, N. (2017). The insular cortex. *Current Biology*, 27(12), R580–R586.
- Graham, F. K. (1979). Distinguishing among orienting, defense, and startle reflexes. In Kimmel, H. D., van Olst, E. H., & Orlebeke, J. F. (Eds.) *The orienting reflex in humans*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp 137–167.
- Graham, F. K., & Clifton, R. K. (1966). Heart-rate change as a component of the orienting response. *Psychological Bulletin*, 65(5), 306–320.
- Gu, X., FitzGerald, T. H. B., & Friston, K. J. (2019). Modeling subjective belief states in computational psychiatry: Interoceptive inference as a candidate framework. *Psychopharmacology*, 236(8), 2405–2412.
- Hahn, W. W. (1973). Attention and heart rate: A critical appraisal of the hypothesis of Lacey and Lacey. *Psychological Bulletin*, 79(1), 59–70.
- Hodges, W. F., & Fox, R. (1965). Effect of arousal and intelligence on binocular rivalry rate. *Perceptual & Motor Skills*, 20, 71–75.
- Hodossy, L., & Tsakiris, M. (2020). Wearing your heart on your screen: Investigating congruency-effects in autonomic responses and their role in interoceptive processing during biofeedback. *Cognition*, 194, 104053.
- Hohwy, J., Roepstorff, A., & Friston, K. J. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, 108(3), 687–701.
- Iatsenko, D., Bernjak, A., Stankovski, T., Shiogai, Y., Owen-Lynch, P. J., Clarkson, P. B. M., ..., Stefanovska, A. (2013). Evolution of cardiorespiratory interactions with age. *Philosophical Transactions of the Royal Society A*, 371(1997), 20110622.
- Iatsenko, D., McClintock, P. V. E., & Stefanovska, A. (2015). Nonlinear mode decomposition: A noise-robust, adaptive decomposition method. *Physical Review E*, 92(3), 032916.
- Iatsenko, D., McClintock, P. V. E., & Stefanovska, A. (2016). Extraction of instantaneous frequencies from ridges in time-frequency representations of signals. *Signal Processing*, 125, 290–303.
- Jennings, J. R. (1986). Bodily changes during attention. In Coles, M. G. H., Donchin, E., & Porges, S. W. (Eds.) *Psychophysiology: systems, processes, and applications*, New York & London: Guilford Press, pp 268–289.
- Kassambara, A. (2020). ggpubr: ‘ggplot2’ based publication ready plots.
- Kaufmann, T., Sütterlin, S., Schulz, S. M., & Vögele, C. (2011). ARTiiFACT: A tool for heart rate artifact processing and heart rate variability analysis. *Behavior Research Methods*, 43(4), 1161–1170.
- Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, J. S., Feinstein, J. S., ..., Paulus, M. P. (2018). Interoception and mental health: A roadmap. *Biological Psychiatry: Cognitive Neuroscience & Neuroimaging*, 3, 501–513.
- Knapen, T., Brascamp, J., Pearson, J., van Ee, R., & Blake, R. (2011). The role of frontal and parietal brain areas in bistable perception. *Journal of Neuroscience*, 31(28), 10293–10301.
- Koch, E. B. (1932). Die irradiation der pressorezeptorischen Kreislaufreflexe. *Klinische Wochenschrift*, 2, 225–227.
- Lacey, B. C., & Lacey, J. I. (1974). Studies of heart rate and other bodily processes in sensorimotor behavior. In Obrist,

- P. A., Black, A. H., Brener, J., & DiCara, L. V. (Eds.) *Cardiovascular psychophysiology: Current issues in response mechanisms, biofeedback and methodology*, Chicago, IL: Aldine Publishing Co., pp 538–564.
- Lacey, J. I. (1959). Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In Rubinstein, E. A., & Parloff, M. B. (Eds.) *Research in psychotherapy*, Washington, DC: American Psychological Association, pp 160–208.
- Lacey, J. I. (1967). Somatic response patterning and stress: Some revisions of activation theory. In Appley, M. H., & Trumbull, R. (Eds.) *Psychological stress: issues in research*, New York, NY: Appleton-Century-Crofts, pp 14–37.
- Lacey, J. I. (1972). Some cardiovascular correlates of sensorimotor behavior: Examples of visceral afferent feedback? In Hockman, C. H. (Ed.) *Limbic system mechanisms and autonomic function*, Springfield, IL: Charles C. Thomas, pp 175–196.
- Lacey, J. I., Kagan, J., Lacey, B. C., & Moss, H. A. (1963). The visceral level: Situational determinants and behavioral correlates of autonomic response patterns. In Knapp, P. H. (Ed.) *Expression of the emotions in man*, New York, NY: International Universities Press, pp 161–196.
- Lacey, J. I., & Lacey, B. C. (1958). The relationship of resting autonomic activity to motor impulsivity. In *Proceedings of the Association for Research in Nervous & Mental Disease*, vol 36, pp 144–209.
- Lacey, J. I., & Lacey, B. C. (1970). Some autonomic-central nervous system interrelationships. In Black, P. (Ed.) *Physiological correlates of emotion*, New York & London: Academic Press, pp 205–227.
- Lenth, R. (2020). emmeans: Estimated marginal means, aka least-squares means.
- Libby, W. L., Lacey, B. C., & Lacey, J. I. (1973). Pupillary and cardiac activity during visual attention. *Psychophysiology*, 10(3), 270–294.
- Limanowski, J. (2017). (Dis-)attending to the body: Action and self-experience in the active inference framework. In Metzinger, T., & Wiese, W. (Eds.) *Philosophy and predictive processing*, pp. 1–13. Frankfurt am Main: MIND Group.
- Lumer, E. D., Friston, K. J., & Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science*, 280(5371), 1930–1934.
- Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago Face Database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47(4), 1122–1135.
- Makowski, D., Sperduti, M., Blondé, P., Nicolas, S., & Piolino, P. (2020). The heart of cognitive control: Cardiac phase modulates processing speed and inhibition. *Psychophysiology*, 57(3), e13490.
- Moran, R. J., Symmonds, M., Dolan, R. J., & Friston, K. J. (2014). The brain ages optimally to model its environment: Evidence from sensory learning over the adult lifespan. *PLoS Computational Biology*, 10(1), e1003422.
- Motyka, P., Grund, M., Forschack, N., Al, E., Villringer, A., & Gaebler, M. (2019). Interactions between cardiac activity and conscious somatosensory perception. *Psychophysiology*, 56(10), e13424.
- Mulcahy, J. S., Larsson, D. E. O., Garfinkel, S. N., & Critchley, H. D. (2019). Heart rate variability as a biomarker in health and affective disorders: A perspective on neuroimaging studies. *NeuroImage*, 202, 116072.
- Nakao, H., Ballim, H. M., & Gellhorn, E. (1956). The role of the sino-aortic receptors in the action of adrenaline, nor-adrenaline and acetylcholine on the cerebral cortex. *Electroencephalography & Clinical Neurophysiology*, 8(3), 413–420.
- Nassar, M. R., Bruckner, R., Gold, J. I., Li, S.-C., Heekeren, H. R., & Eppinger, B. (2016). Age differences in learning emerge from an insufficient representation of uncertainty in older adults. *Nature Communications*, 7, 11609.
- Obrist, P. A. (1963). Cardiovascular differentiation of sensory stimuli. *Psychosomatic Medicine*, 25, 450–459.
- Ottaviani, C. (2018). Brain-heart interaction in perseverative cognition. *Psychophysiology*, 55(7), e13082.
- Owens, A. P., Allen, M., Ondobaka, S., & Friston, K. J. (2018). Interoceptive inference: From computational neuroscience to clinic. *Neuroscience & Biobehavioral Reviews*, 90, 174–183.
- Park, G., & Thayer, J. F. (2014). From the heart to the mind: Cardiac vagal tone modulates top-down and bottom-up visual perception and attention to emotional stimuli. *Frontiers in Psychology*, 5, 278.
- Park, H.-D., Correia, S., Ducorps, A., & Tallon-Baudry, C. (2014). Spontaneous fluctuations in neural responses to heartbeats predict visual detection. *Nature Neuroscience*, 17(4), 612–618.
- Park, H.-D., & Tallon-Baudry, C. (2014). The neural subjective frame: From bodily signals to perceptual consciousness. *Philosophical Transactions of the Royal Society B*, 369(20130208), 1–9.
- Parr, T. (2020). Inferring what to do (and what not to). *Entropy*, 22, 536.
- Parr, T., Corcoran, A. W., Friston, K. J., & Hohwy, J. (2019). Perceptual awareness and active inference. *Neuroscience of Consciousness*, 5(1), niz012.
- Parr, T., & Friston, K. J. (2017). Uncertainty, epistemics and active inference. *Journal of the Royal Society Interface*, 14(20170376), 1–10.
- Perrykard, K., & Hohwy, J. (2020). Fidgeting as self-evidencing: A predictive processing account of non-goal-directed action. *New Ideas in Psychology*, 56(100750), 1–8.
- Peters, A., McEwen, B. S., & Friston, K. J. (2017). Uncertainty and stress: Why it causes diseases and how it is mastered by the brain. *Progress in Neurobiology*, 156, 164–188.
- Petzschner, F. H., Weber, L. A. E., Gard, T., & Stephan, K. E. (2017). Computational psychosomatics and computational psychiatry: Toward a joint framework for differential diagnosis. *Biological Psychiatry*, 82, 421–430.
- Pezzulo, G., Rigoli, F., & Friston, K. J. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, 134, 17–35.
- Porges, S. W. (1992). Autonomic regulation and attention. In Campbell, B. A., Hayne, H., & Richardson, R. (Eds.) *Attention and information processing in infants and adults: Perspectives from human and animal research*, pp. 201–223. New York and London: Psychology Press.
- Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. A polyvagal theory. *Psychophysiology*, 32(4), 301–318.
- Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology*, 74(2), 116–143.
- Porges, S. W., & Raskin, D. C. (1969). Respiratory and heart rate components of attention. *Journal of Experimental Psychology*, 81(3), 497–503.
- Quadt, L., Critchley, H. D., & Garfinkel, S. N. (2018). The neurobiology of interoception in health and disease. *Annals of the New York Academy of Sciences*, 1428(1), 112–128.
- Quigley, K. S., & Berntson, G. G. (1996). Autonomic interactions and chronotropic control of the heart: Heart period versus heart rate. *Psychophysiology*, 33(5), 605–611.
- R Core Team (2019). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

- Rau, H., & Elbert, T. (2001). Psychophysiology of arterial baroreceptors and the etiology of hypertension. *Biological Psychology*, 57(1–3), 179–201.
- Rau, H., Pauli, P., Brody, S., Elbert, T., & Birbaumer, N. (1993). Baroreceptor stimulation alters cortical activity. *Psychophysiology*, 30(3), 322–325.
- Ribeiro, M. J., & Castelo-Branco, M. (2019). Neural correlates of anticipatory cardiac deceleration and its association with the speed of perceptual decision-making, in young and older adults. *NeuroImage*, 199, 521–533.
- RStudio Team. (2015). *Rstudio: Integrated development for R*. Boston, MA: RStudio, Inc.
- Salomon, R., Ronchi, R., Dönn, J., Bello-Ruiz, J., Herbelin, B., Martet, R., ..., Blanke, O. (2016). The insula mediates access to awareness of visual stimuli presented synchronously to the heartbeat. *Journal of Neuroscience*, 36(18), 5115–5127.
- Saper, C. B., & Loewy, A. D. (1980). Efferent connections of the parabrachial nucleus in the rat. *Brain Research*, 197(2), 291–317.
- Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B*, 371(1708), 1–10.
- Seth, A. K., & Tsakiris, M. (2018). Being a beast machine: The somatic basis of selfhood. *Trends in Cognitive Sciences*, 22(11), 969–981.
- Shannon, R. W., Patrick, C. J., Jiang, Y., Bernat, E., & He, S. (2011). Genes contribute to the switching dynamics of bistable perception. *Journal of Vision*, 11(3), 8.
- Shoemaker, J. K., Wong, S. W., & Cechetto, D. F. (2012). Cortical circuitry associated with reflex cardiovascular control in humans: Does the cortical autonomic network “speak” or “listen” during cardiovascular arousal. *Anatomical Record*, 295(9), 1375–1384.
- Smith, R., Thayer, J. F., Khalsa, S. S., & Lane, R. D. (2017). The hierarchical basis of neurovisceral integration. *Neuroscience & Biobehavioral Reviews*, 75, 274–296.
- Sokolov, E. N. (1960). Neural models and the orienting reflex. In Brazier, M. A. B. (Ed.) *The central nervous system and behavior*, pp. 187–276. New York NY: Josiah Macy Jr Foundation.
- Sokolov, E. N. (1969). The modeling properties of the nervous system. In Cole, M., & Maltzman, I. (Eds.) *A handbook of contemporary Soviet psychology*, pp. 671–704. New York, NY: Basic Books.
- Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A. E., Paliwal, S., Gard, T., ..., Petzschner, F. H. (2016). Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience*, 10(550), 1–27.
- Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders*, 61(3), 201–216.
- Thayer, J. F., & Lane, R. D. (2009). Claude Bernard and the heart–brain connection: Further elaboration of a model of neurovisceral integration. *Neuroscience & Biobehavioral Reviews*, 33(2), 81–88.
- Uddin, L. Q., Nomi, J. S., Hébert-Seropian, B., Ghaziri, J., & Boucher, O. (2017). Structure and function of the human insula. *Journal of Clinical Neurophysiology*, 34(4), 300–306.
- van Rij, J., Wieling, M., Baayen, H. R., & van Rijn, H. (2017). itsadug: Interpreting time series and autocorrelated data using GAMMs.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ..., Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.
- Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The shine toolbox. *Behavior Research Methods*, 42(3), 671–684.
- Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 65, 95–114.
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 73(1), 3–36.
- Wood, S. N. (2017). *Generalized additive models: An introduction with R. Texts in Statistical Science*, (2nd ed.). Boca Raton, FL: CRC Press.
- Zhang, Z., & Oppenheimer, S. M. (1997). Characterization, distribution and lateralization of baroreceptor-related neurons in the rat insular cortex. *Brain Research*, 760(1–2), 243–50.
- Zhang, Z. H., Dougherty, P. M., & Oppenheimer, S. M. (1998). Characterization of baroreceptor-related neurons in the monkey insular cortex. *Brain Research*, 796(1–2), 303–306.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

6

Restless hearts and wandering minds: The cardiac correlates of task-unrelated thought

The findings of the previous chapter exemplify how the conceptual resources of the active inference framework can be used to inform psychophysiological studies in ways that deliver new insights about old phenomena, like attentional bradycardia and cardiac orienting. What’s more, this work demonstrates how the adoption of an active inference perspective can help to unveil a surprisingly active component of binocular rivalry – a phenomenon that might otherwise seem entirely passive and fundamentally perceptual in character.

If the interpretation presented in Chapter 5 is on the right track, then the cardiac domain affords an opportunity for prediction error minimisation that is increasingly exploited in the context of accumulating uncertainty. Together with Parr and colleagues’ (2019) computational model of binocular rivalry, cardiac modulation might be viewed as another expression of covert action – one that parallels and complements the mental actions driving bistable perception. The picture that is thus beginning to emerge is one of a brain that presides over multiple modalities in an orchestrated effort to assuage uncertainty.

The investigation of covert forms of active inference promises to open up exciting new ways to think about the ligatures that bind brain, body, and world. However, the

perceptual dynamics focused on thus far play an admittedly limited role in the broader scheme of biological regulation. Ultimately, the reduction of uncertainty over sensory states is only functional to the extent that it helps to optimise inferences about what to *do*. Insofar as perception is in the service of action, these sorts of uncertainty-resolving dynamics figure as part of an essential, but nonetheless precursory moment within the broader sweep of the agent’s adaptive arc.

With this caveat in mind, the study reported in this chapter sought to make inroads towards a more explicitly action-oriented perspective on brain-heart communication. The goal of this experiment was to investigate the linkage between covert (mental and physiological) activity and overt motor behaviour. As was the case in Chapter 5, this study inherits from a rich tradition of psychophysiological research involving a number of overlapping constructs, including attention, inhibition, and cognitive (or executive) control. In order to make some of the continuities between this work and the previous chapter explicit, the historical development of this literature is surveyed below.

6.0.1 Cardiac deceleration: From orientation to action

As briefly alluded to in the Discussion of Chapter 5, there is a strong affinity between the Laceys’ early work on cardiac deceleration and the cardiac component of the orienting response. This affinity was first highlighted in a seminal review by Graham and Clifton (1966), who pointed out that the Lacey’s *intake-rejection hypothesis* (Lacey, 1959; Lacey et al., 1963) and contemporaneous work on the orienting reflex in the Soviet Union (Sokolov, 1960, 1963b) both ascribe a functional role to autonomic feedback in the enhancement of sensory receptivity. The major contribution of this review, however, was to resolve an apparent contradiction between the two bodies of work: While Sokolov (1963a) had inferred that cardiac orienting manifests as a *speeding* rather than slowing of the heart rhythm, Graham and Clifton (1966) convincingly argued that this interpretation had likely conflated orienting with the defense or startle reflex. Subsequent research confirmed the hypothesis that cardiac orienting is characterised by a dominant deceleration response (Graham, 1979).

As interest in the orienting reflex spread from the Soviet Union to North America, and as evidence of the autonomic nervous system’s sensitivity to distinct sensory properties accrued through the work of the Laceys and others, scholars began to consider the possible implications of such phenomena for overt behaviour. Following Kvasov (1965), Germana (1969) proposed that Pavlov’s (1927) original characterisation of the orienting response as the “*What is it?*” reflex might be more fittingly recast in terms of the question: “*What’s to be done?*” (see also Berlyne 1960; Pribram and McGuinness

1975). Perspectives such as these, which shift the emphasis from perceptual recognition to response preparation, are of course entirely in keeping with an active inference-style gloss on sensorimotor coupling (cf. Pezzulo and Cisek 2016; Seth 2014).

The tight coupling between perception and action was not lost on the Laceys, who consistently referred to the effects of visceral afferent feedback on *sensorimotor* (rather than purely perceptual) processing (e.g., Lacey and Lacey 1958; Lacey 1967; Lacey and Lacey 1974). Indeed, as the quote featured in the Introduction of the last chapter succinctly articulates, the Laceys construed cardiac deceleration as a mechanism that not only enhances “the organism’s receptivity to afferent stimulation,” but also its “readiness to make effective responses to such stimulation” (Lacey 1972, p. 183). The Janus-faced character of this formulation was supported by a series of signalled (i.e. cued) reaction time experiments, in which a warning signal informs the participant about the impending occurrence of the imperative stimulus to which they must respond (Lacey, 1967; Lacey and Lacey, 1970, 1974). These studies revealed two key findings: (1) the heart beat progressively slows over the course of the foreperiod (i.e. preparatory interval) leading up to the imperative stimulus; and (2) the depth of such anticipatory deceleration correlated (albeit modestly) with reaction time, whereby greater deceleration predicted faster responding (a surprising result given the widespread assumption that heart rate acceleration indexes behavioural arousal).

The Laceys interpreted their findings (and corroborative reports from independent labs; e.g., Connor and Lang 1969; Obrist et al. 1969; Webb and Obrist 1970; cf. Duncan-Johnson and Coles 1974; Porges 1972; Jennings et al. 1970; Nowlin et al. 1970) as offering further support for their visceral afferent feedback hypothesis, inferring that decreased stimulation of the arterial baroreceptors promoted “increased receptivity to external stimulation and increased ease of effective motor response” (Lacey and Lacey 1974, p. 548). In much the same way as cardiac slowing during sustained attention to a visual stimulus or verbal narrative was supposed to optimise the organism’s capacity to receive and process sensory information (Lacey et al., 1963; Lacey, 1967), anticipatory deceleration was posited to optimise the system’s capacity to register and respond to the impending signal as rapidly as possible. One might construe this phenomenon as an anticipatory analogue of the standard orienting response, which is classically evoked by an unexpected change in the environment.¹

¹For completeness, it’s worth noting that the cardiac response classically evoked during the foreperiod preceding the imperative stimulus typically evinces a triphasic form, consisting of an initial, small deceleration (most likely an orienting response to the warning signal), a subsequent accelerative component, and finally a second, more pronounced deceleration (Bohlin and Kjellberg, 1979). Interestingly, the magnitude of the acceleration phase is sensitive to stimulus properties and response demands (see, e.g., Coles and Duncan-Johnson 1975; Higgins 1971; Lang et al. 1978).

As the Laceys (1974) themselves acknowledged, the signalled reaction time paradigm confounds the ‘intention to note and detect’ stimuli with the ‘intention to respond’ to them. Consequently, it cannot be inferred that the cardiac deceleration observed in the foreperiod preceding the imperative stimulus was driven by expectant attentional engagement with one’s environment *per se*. However, this concern is easily allayed by evidence of cardiac slowing in anticipation of delayed performance feedback (De Pascalis et al., 1995; Lacey and Lacey, 1970), and in yoked reaction time conditions in which participants simply observed task stimuli without overtly responding to them (Lang et al., 1978). Such studies provide unequivocal proof that anticipatory cardiac deceleration can be elicited independently of motor preparation or associated processes (i.e. in a similar vein to how the findings reported in Chapter 5 revealed graded differences in cardiac responses despite invariant task demands).

A more enduring line of criticism of the Laceys’ conclusions argued that the cardiac slowing associated with orienting and sustained attention might simply be a consequence of a more general inhibitory response, given that attending to one’s environment calls for motor quiescence (Obrist 1968, see also Kahneman 1973). While this view is compatible in some respects with that of the Laceys (after all, both involve forms of inhibition designed to improve sensory intake), the crucial point of disagreement here concerns whether cardiac slowing constitutes an instrument of attentional regulation, or merely a secondary, epiphenomenal consequence of it. The latter view, which was expressed most vigorously in the form of the cardiac-somatic coupling hypothesis (Obrist and Webb, 1967; Obrist et al., 1970; Obrist, 1976, 1981), would deny that visceral afferent feedback plays any significant role in determining cognitive and behavioural function.

We need not retrace the various disputes that unfolded between subscribers of these views in the 1970s and ’80s; debates which, as Porges (1992) laments, did much to obscure their substantial communality. Both perspectives are in basic agreement about the allostatic character of cardiac regulation, and the tight linkage between cardiovascular control and cognitive-behavioural states. Of greater pertinence here is the basic insight that attending to one’s environment and preparing a particular course of action might both be facilitated by the cessation (or dampening down) of ongoing activity (Obrist et al., 1970). This focus on motor inhibition, and its connection with cardiac regulation, would resurface in a series of psychophysiological studies of executive function conducted in the 1990s and 2000s.

6.0.2 Selective attention, inhibition, and executive control

In order to effectively address the question “*What’s to be done?*”, the organism must be capable of inferring the best available policy at its disposal, and of enacting that policy. As intimated in previous chapters, action selection entails the reduction of uncertainty about the likely outcomes of alternative policies. One can construe ‘attentional’ acts – be they covert, as in binocular rivalry, or overt, as when orienting toward a sound – as simple exploratory policies designed to garner information about the environment that can be exploited in the service of future behaviour. Having thus updated one’s beliefs about the state of the environment, one is in a position to opt for the policy most likely to solicit preferred outcomes (cf. [Limanowski and Friston 2018](#)).

The silent partner in all this is inhibition: One does not simply infer what to do, but also what *not* to (at least, not yet). By analogy to the selective sampling or gating of sensory inputs, psychophysiological accounts of attention sometimes include action selection and inhibition within their purview (even if the mechanisms mediating such processes are supposed to differ from those involved in stimulus processing; e.g., [Kahneman 1973](#); [Posner et al. 2007](#); cf. the concept of ‘supervisory attention’, [Norman and Shallice 1986](#)). Others consider attention and inhibitory control to be closely related, potentially overlapping modes of executive function ([Diamond, 2013](#)) or cognitive control (e.g., [Jennings 1986](#)). Although such taxonomic minutiae and their attendant boundary disputes are not of concern here, active inference clearly favours a broader, unifying interpretation of attention in both perception and action ([Brown et al., 2011](#)).

Given the tight relation between sensory receptivity and response preparation as conceived in both the orienting reflex (e.g., [Sokolov 1963b](#), p. 118) and the intake-rejection hypothesis, as well as the closely-related issue of motor inhibition that features prominently in the cardiac-somatic coupling hypothesis, surprisingly little contemporaneous psychophysiological research attempted to isolate the effects of action preparation on cardiac dynamics.

Early work using Stroop’s (1935) Color-Word Interference Test revealed a decline in average heart rate under response conflict (i.e. when competing, incompatible responses are elicited), with more challenging conditions evoking stronger rate reductions ([Elliott, 1969](#); [Elliott et al., 1970](#)). Coles and Duncan-Johnson (1975) reported that phasic deceleration responses were only evoked during the preparatory interval on trials in which participants had been cued to respond to a predictable imperative stimulus, as opposed to simply observing it (see also [van der Molen et al. 1983](#)). These findings imply that response preparation is a sufficient condition for the elicitation of anticipatory cardiac

deceleration, and moreover, that cardiac deceleration is sensitive to conflicting response demands.

More recent investigation of the psychophysiological correlates of motor preparation and inhibition have revealed the autonomic adjustment of heartbeat timing to be “an integral component of the brain’s control of moment-to-moment action regulation” (Jennings and van der Molen 2002, p. 346). Interestingly, this body of work supplements the well-established association between graded cardiac slowing and temporal expectation with evidence of a more nuanced relation implicating cardiac regulation in the mental representation of possible actions and task sets. The subtle but important shift from motor activation to mental preparation, based in part on evidence that preparatory processes operate over central representations rather than motor efferent processes (see Jennings and van der Molen 2005), sits comfortably with the active inference account of both mental action as top-down precision modulation, and motor action as a consequence of inference on counterfactual representations (see Chapter 3).

An important methodological step in the development of more refined psychophysiological models of inhibitory control was the progression beyond simple reaction time tasks. The fixed foreperiod paradigm employed by the Lacey and others probes a form of response preparation that loads heavily on temporal prediction (Niemi and Näätänen, 1981); the participant is poised to issue a prescribed response, and can exploit the temporal regularity between cue and imperative stimulus to reduce uncertainty about the appropriate time to do so (see Bohlin and Kjellberg 1979; van der Molen et al. 1987). Paradigms such as the Go/NoGo (Donders, 1969) and stop-signal tasks (Lapin and Eriksen, 1966; Logan and Cowan, 1984) complicate this situation by requiring participants to suppress or countermand a prepotent response tendency on a subset of trials, thus engendering uncertainty over action execution (Braver et al., 2001). Such paradigms furnish more informative data about the evolution of inhibitory processes (and their failures) over time.

An early experiment using the Go/NoGo paradigm revealed that NoGo stimuli (which require participants to refrain from responding) were followed by longer inter-beat intervals than Go stimuli (which require a simple or choice-discrimination response), indicating that recovery from anticipatory deceleration was delayed in the NoGo condition (van der Molen et al. 1989; see also van der Molen et al. 1983; van der Veen et al. 2000). Subsequent work analysing cardiac responses as a function of NoGo trial performance revealed that erroneous responses tend to elicit the curtailment of cardiac deceleration following stimulus onset, while cardiac slowing persists into the post-stimulus inter-beat interval when responses are successfully inhibited (van Boxtel et al., 2001). A similar pattern of protracted cardiac slowing was also observed during complete or partial

response inhibition on a stop-signal variant of the task (cf. [Jennings et al. 1992](#)). Jennings and van der Molen ([2002](#)) interpret such data as evidence that cardiac deceleration indexes the inhibition of competing action representations during preparatory intervals.

6.0.3 Performance monitoring and error processing

In addition to teasing apart differences between temporal expectation and response preparation in the lead up to a stimulus event, paradigms such as the Go/NoGo task also afford opportunities to investigate psychophysiological correlates of performance monitoring and error (or conflict) processing. Performance monitoring is a form of cognitive control concerned with the detection of unexpected action-outcome contingencies, and the adaptive modulation of behaviour in light of such occurrences (for reviews, see [Alexander and Brown 2010](#); [Desmet et al. 2011](#); [Ullsperger et al. 2014](#)). In the context of a Go/NoGo task, this might manifest as the recognition that one failed to inhibit a prepotent Go response on a NoGo trial, prompting the adoption of a slower, more careful response strategy on subsequent trials (i.e. *post-error slowing*; [Laming 1979](#); [Rabbitt 1966](#)).

Performance monitoring requires the agent to track the consequences of its actions in the context of (potentially competing) goal states and (potentially changing) environmental conditions. It thus entails a kind of online feedback processing in which observed states are evaluated in relation to expected outcomes, such that motor action can be flexibly modified or reconfigured in accordance with unfolding events ([Botvinick et al., 2001](#); [Holroyd and Coles, 2002](#); [Ridderinkhof et al., 2004](#); [Ullsperger and von Cramon, 2004](#)). This formulation is of course entirely at home with cybernetic- and active inference-inspired accounts of adaptive behaviour, whereby remedial actions are driven by (prediction) error minimisation. Of particular relevance here are recent attempts to incorporate psychophysiological perspectives on attentional orienting and inhibitory control within performance monitoring/error processing frameworks ([Notebaert et al., 2009](#); [Wessel, 2018](#)).

An early study by Danev and de Winter ([1971](#)) reported transient cardiac deceleration following visual feedback informing the participant that they had responded incorrectly or too slowly on a choice reaction time task. The authors speculated that this phenomenon might reflect attentional orienting to external stimuli, under the assumption that error feedback constitutes a more salient or informative signal than confirmation of a correct response. More recent reports have consistently documented cardiac slowing in response to performance feedback, with negative feedback tending to evoke deeper, more protracted deceleration than positive ([Crone et al., 2003, 2004, 2005](#); [Fraga González](#)

et al., 2019; Groen et al., 2007; Herman et al., 2021; Kastner et al., 2017; Somsen et al., 2000; van der Veen et al., 2004a,b).

Cardiac deceleration has also been observed in the wake of performance errors during tasks that do not furnish explicit feedback (Bastin et al., 2017; Fiehler et al., 2004; Hajcak et al., 2003; Spruit et al., 2018). Interestingly, the magnitude of the cardiac response has been reported to vary with error awareness: Perceived errors are associated with an enhanced deceleration response compared to errors that go unnoticed (Wessel et al., 2011). Relatedly, a recent study observed that post-error slowing following threshold-level visual stimulation predicted more accurate discrimination performance and higher visibility ratings; moreover, these effects were disrupted by the provision of false cardiac feedback (Lukowska et al., 2018). Findings such as these raise the intriguing possibility that cardiac rate changes not only index error-related attentional processing, but might also contribute to the awareness and evaluation of one's performance (cf. Bury et al. 2019; Damasio 1996; Hajcak et al. 2003; Skora et al. 2021).

The modulation of cardiac dynamics following erroneous task responses would seem to complicate the interpretation of differential post-stimulus cardiac deceleration as an index of inhibitory control. It has long been known that individuals are typically sensitive to their errors on choice response tasks in the absence of explicit feedback (Rabbitt, 1968), an observation that seems to hold for at least some varieties of the Go/NoGo task (Head and Helton, 2013). It is also notable that many such tasks engender prepotent response biases by assigning NoGo/Stop stimuli to a relatively small proportion of trials. Such events may therefore promote attentional orienting not only by virtue of their salience from a performance monitoring perspective, but also by dint of their relative infrequency (see Braver et al. 2001; Notebaert et al. 2009).

In sum, although the impressive body of work by Jennings, van der Molen, and colleagues would seem to substantiate the claim that cardiac deceleration indexes the deployment of inhibitory control processes in the context of action preparation, the emergent relation between cardiac activity and performance monitoring reasserts the difficulty of teasing apart different components of executive control as expressed in heartbeat dynamics. Indeed, as Jennings and van der Molen would themselves later concede, the primary role of inhibitory processing in their scheme manifests “only as part of a more general attentional response with components of alerting and orienting” (Jennings et al. 2009, p. 1176). It might therefore prove more productive to construe top-down fluctuations in cardiac dynamics as the consequence of multiple interacting control processes united by the common imperative to optimise sensorimotor integration with the environment – a view in keeping with the spirit of the Lacey's intake-rejection hypothesis, and with the theoretical perspective advanced in Chapter 5.

6.0.4 The present manuscript

While task-related fluctuations in cardiac dynamics might not map neatly onto different components of executive function, it might still be possible to differentiate systematic patterns of peristimulus cardiac activity under differing states of attentional or executive control. A novel way to approach this question is through the lens of *mind-wandering*, a pervasive psychological phenomenon involving the spontaneous fluctuation of attentional states over time. By adopting methodological techniques from mind-wandering research, one can perform a quasi-experimental analysis that compares behavioural task performance and accompanying physiological measures as a function of attentional state. In this way, my co-authors and I were able to investigate how cardiac activity during attentive observation, motor inhibition, and performance monitoring varies across distinct classes of mental action (namely, whether attention was directed on-task or not).

Perhaps somewhat surprisingly given the extensive body of psychophysiological research on the cardiac correlates of attention, very few mind-wandering studies have attempted to quantify the relation between cardiac activity and attentional dynamics. Of the few reports that exist (reviewed in the following Introduction), analysis is almost entirely limited to heart rate and variability metrics computed over 30 s epochs. There appears to be no data in the literature on the way cardiac states evolve over shorter (e.g. beat-by-beat) or longer time frames (e.g., minutes), despite the knowledge that attentional states may fluctuate rapidly, and are subject to substantial time-on-task effects. The following experiment thus sought to furnish the first in-depth study of the cardiac correlates of mind-wandering across multiple temporal scales.

Restless hearts and wandering minds: The cardiac correlates of task-unrelated thought

Andrew W. Corcoran, Thomas Andrillon, Jakob Hohwy

Abstract

Mind-wandering is a ubiquitous psychological phenomenon characterised by the spontaneous disengagement of attention from goal-directed behaviour. Although cardiac activity is reliably modulated by attentional orienting and vigilance, little is known about the behaviour of the heart rhythm during episodes of mind-wandering. Here, we report behavioural and electrocardiogram data from 23 adults during performance of a modified Sustained Attention to Response Task (SART). Inter-beat interval and heart rate variability estimates increased linearly over the course of the task, in correspondence with increased rates of task-unrelated thought and behavioural performance decrements. Cardiac estimates also accounted for unexplained variance in reaction times on SART trials. Cardiac measures did not significantly differ between attentional states; however, our analysis revealed evidence of non-uniform stimulus and response onset timing across the cardiac cycle during periods of task-unrelated thought. These data suggest the wandering mind may help synchronise action with the cardiac cycle, which may in turn adapt its timing in accordance with rhythmic sensory stimulation.

Keywords: Mind-wandering; Attention; Vigilance; Inter-beat interval; Heart rate variability; Time-on-task

6.1 Introduction

The ability to endogenously sustain one’s attention on the environment over a protracted period of time is an essential component of adaptive behaviour. Lapses of attention during the performance of routine activities may engender detrimental consequences ranging from the merely inconvenient to the outright catastrophic ([Reason, 1990](#)); more chronic difficulties maintaining attention on a given object or task may lead to significant debilitation, secondary impairment, and socio-economic disadvantage. It is understandable, then, that transient bouts of inattention have traditionally been conceived in terms of deficiency or failure (e.g., [Brown 1927](#)): an unfortunate limitation of capacity or resources; an undesirable weakness of will or constitution.

Although considerable research effort has been devoted towards the reduction (or mitigation) of attentional lapses during task performance, a growing contingent of scholars have begun to treat distraction and disengagement as objects of scientific inquiry in their own right. Several theoretical accounts have recently been proposed that highlight the broad spectrum of cognitive processes and conscious states that may come to the fore when one’s attentional focus on the immediate environment begins to wane, including some of the adaptive functions these processes might serve (e.g., [Andrillon et al. 2019](#); [Christoff et al. 2016](#); [Irving 2016](#); [Metzinger 2013](#); [Mildner and Tamir 2019](#); [Mittner et al. 2016](#); [Seli et al. 2018](#); [Smallwood and Schooler 2015](#)). This heterogeneous cluster of phenomena, encompassing such disparate mental activities as daydreaming, planning, and rumination, is most commonly grouped under the rubric of *mind-wandering*.

6.1.1 Mind-wandering methodology

The scientific investigation of mind-wandering is complicated not only by the unobservable nature of thought in general, but also by the characteristically spontaneous, task- or stimulus-independent nature of mind-wandering in particular. Most studies of mind-wandering address this issue by employing some form of thought sampling technique, in which participants report the content of their thoughts either when they endogenously realise that their stream of consciousness has been diverted off-task, or in response to exogenous probes that interrupt ongoing activity ([Smallwood and Schooler, 2006](#)).

Under laboratory conditions, thought sampling techniques are typically combined with experimental paradigms designed to render objective measures of attentional fluctuation. Such tasks often exploit insights derived from decades of attention and vigilance research to construct protocols that are conducive to mind-wandering (see, e.g., [Antrobus et al. 1966](#); [Giambra 1995](#)).

One widely-used method for engendering and quantifying the behavioural correlates of mind-wandering is the Sustained Attention to Response Task (SART; [Robertson et al. 1997](#)). The SART is a variety of Go/NoGo task in which a pre-specified target stimulus (e.g., the number “3”) must be discriminated from a series of non-target stimuli or foils (e.g., other digits). In a departure from Continuous Performance Task variants of the Go/NoGo paradigm classically employed in vigilance research ([Rosvold et al., 1956](#)), the SART requires participants to respond to each occurrence of a non-target stimulus, while withholding responses to target stimuli. These response instructions, coupled with the frequent, rhythmic presentation of Go stimuli, establish a prepotent response bias that must be intermittently suppressed or overridden in order to prevent automatic responding on NoGo trials ([Cheyne et al., 2006, 2011](#); [Manly et al., 1999](#); [Robertson et al., 1997](#)).

Transient lapses of attention during SART performance manifest as the failure to inhibit responses on NoGo trials (i.e. errors of commission), the failure to issue a response on Go trials (error of omission), and the speeding of reaction times on Go trials preceding such errors (as compared to trials preceding successfully inhibited responses; [Cheyne et al. 2009](#); [Manly et al. 1999, 2000](#); [Robertson et al. 1997](#)). These behavioural indices show convergent validity with self-report inventories of attentional lapses ([Cheyne et al., 2006](#); [Farrin et al., 2003](#); [Robertson et al., 1997](#); [Smilek et al., 2010a](#)), and are associated with increased rates of mind-wandering as assessed via thought sampling (e.g., [Christoff et al. 2009](#); [McVay and Kane 2009](#); [Smallwood et al. 2004a](#)). The SART has also been used to characterise attentional or executive control deficits in a variety of clinical populations, including those affected by traumatic brain injury ([Dockree et al., 2004](#); [Manly et al., 2003](#); [O’Keeffe et al., 2004](#); [Robertson et al., 1997](#)), attention deficit hyperactivity disorder ([Bellgrove et al., 2006, 2005](#); [Johnson et al., 2007a,b](#); [Manly et al., 2001](#)), affective disorder ([Farrin et al., 2003](#); [Smallwood et al., 2007](#)), schizophrenia ([Chan et al., 2009](#)), and work stress ([van der Linden et al., 2005](#)).

6.1.2 The psychophysiology of mind-wandering

In addition to behavioural paradigms such as the SART, mind-wandering researchers have also availed themselves of psychophysiological measures of attentional processing. Chief amongst these has been the use of eye tracking to investigate the ebb and flow of attention during reading ([Faber et al., 2018](#); [Foulsham et al., 2013](#); [Reichle et al., 2010](#); [Smilek et al., 2010b](#); [Uzzaman and Joordens, 2011](#)), computerised learning ([Hutt et al., 2019](#)), and driving ([He et al., 2011](#)). Electrodermal activity/galvanic skin response has also revealed evidence of sensitivity to mind-wandering during the performance on lab-based paradigms ([Blanchard et al., 2014](#)), although results have been somewhat mixed

([Smallwood et al., 2004a,b, 2007](#)). Our prime interest here, however, is in the potential utility of cardiac signals as correlates of attentional states. Such measures have thus far gained little traction in the mind-wandering literature, which is surprising given their prominence in the psychophysiology of attention (discussed in the next section).

To our knowledge, the first studies to analyse cardiac measures in conjunction with thought probes were reported by Smallwood and colleagues ([2004a; 2004b](#)). Consistent with their hypothesis that the content of task-unrelated thought tends to be more physiologically arousing on account of its greater personal significance, Smallwood and colleagues ([2004b](#)) observed a positive correlation between mean heart rate and the frequency of mind-wandering episodes. Similarly, small but significant increases in mean heart rate were reported for periods of task-unrelated thought during the SART ([Smallwood et al., 2004a](#)) and a word-shadowing task ([Smallwood et al., 2007](#)).

Subsequent studies by Ottaviani and colleagues ([Ottaviani et al., 2013, 2015a,b](#)) reported differences in heart rate variability (HRV) as a function of attentional state. Specifically, HRV was found to be significantly reduced during episodes of rumination or worry (‘perseverative cognition’); however, no significant differences in heart rate or HRV were noted between non-perseverative forms of mind-wandering and on-task states. While perseverative cognition was also associated with higher average heart rate during the performance of a laboratory-based tracking/vigilance task ([Ottaviani et al., 2013](#)), this effect was not observed in ambulatory data collected during a 24 hr experience sampling protocol ([Ottaviani et al., 2015a,b](#)).

Finally, some progress has been made towards the development of detection algorithms that classify episodes of mind-wandering on the basis of physiological signals including cardiac features (e.g., [Cheetham et al. 2016; Pham and Wang 2015](#)). While such reports suggest that certain components of the cardiac time series may be predictive of attentional state fluctuation, they do not shed light on the precise nature of any such relation. It isn’t clear, for instance, whether the success of such algorithms is driven by sustained differences in heart rate across task-focused and task-unrelated thought, phasic events such as orienting-like responses associated with attentional switching, onset of meta-awareness that one has been off-task, or some other source of variance.

6.1.3 The psychophysiology of attention

While studies of cardiovascular dynamics in the context of mind-wandering are scarce, the few results highlighted above are broadly consistent with a large corpus of psychophysiological work investigating the cardiac correlates of attentional regulation. Numerous studies dating back to the 1960s documented sustained heart rate modulation

depending on whether participants directed their attention outward towards their sensory environment (associated with heart rate deceleration), or inward towards mental activity or problem solving (associated with heart rate acceleration; [Kagan and Lewis 1965](#); [Lacey et al. 1963](#); [Lacey 1967](#); [Obrist 1963](#)). As reported by Smallwood and colleagues ([2004a](#); [2004b](#)), these effects are typically independent of electrodermal activity (cf. ‘directional fractionation’; [Lacey 1959](#)). Viewed from this perspective, cardiac rate changes might furnish a more general index of task-engagement vs. disengagement (rather than mind-wandering *per se*), insofar as they seem to tap the depth or strength of one’s attentional coupling to the external environment.

Short-term fluctuations in the timing of successive heartbeats (as indexed by the high-frequency component of HRV, or estimates of respiratory sinus arrhythmia) have also been intensively studied in the context of attention. HRV is typically suppressed under conditions of effortful executive regulation such as those called for by vigilance tasks ([Lacey, 1967](#); [Obrist, 1963](#); [Porges and Raskin, 1969](#); [Porges, 1992](#)). Recent work has further suggested that the suppression of HRV during sustained attention may work synergistically with cardiac deceleration, serving to augment sensory processing under conditions of perceptual uncertainty ([Corcoran et al., 2021](#)). While Ottaviani and colleagues’ findings provide little evidence of systematic HRV modulation across task-focused and (non-perseverative) mind-wandering states, further investigation is merited in the interests of contextualising such states within the broader scheme of attentional psychophysiology.

6.1.4 The current study

In contrast to the notion that differences in cardiac activity between task-focused and task-unrelated thought reflect generic fluctuations in arousal states, we adopt a perspective that conceives of cardiac regulation as a form of covert adaptive action. This view is licenced by recent work in computational neuroscience linking beat-by-beat adjustments of cardiac activity to the optimisation of perceptual uncertainty ([Allen et al., 2019](#)). It is also supported by a broad base of psychophysiological evidence demonstrating the exquisite sensitivity of evolving cardiac dynamics to sensorimotor states and performance monitoring (see, e.g., [Lacey and Lacey 1980](#); [Jennings and van der Molen 2005](#); [Öhman et al. 2000](#); [Ullsperger et al. 2014](#); [van der Molen et al. 1985](#)).

One interesting upshot of this perspective is the potential insights cardiac measures might be able to contribute to ongoing debates in the mind-wandering literature. As

already mentioned, cardiac regulation appears to afford a promising index of environmental coupling vs. disengagement – a prominent line of thought in classic and contemporary accounts of mind-wandering (Smallwood and Schooler, 2006). Returning to the SART, there has been considerable disagreement concerning the extent to which this paradigm renders valid indices of attentional regulation, as opposed to inhibitory or motor control processes (see, e.g., Dang et al. 2018; Dillard et al. 2014; Head and Helton 2013; Helton 2009; Manly et al. 2000; Peebles and Bothell 2004; Seli 2016). Insofar as measures of beat-to-beat cardiac fluctuation afford an additional source of data pertaining to (covert) action states, they might help to inform and arbitrate such issues.

This study set out to provide the first in-depth analysis of the cardiac dynamics associated with task-focused and task-unrelated thought as they unfold across multiple timescales. Participants performed two variants of a modified SART over a period of approximately 90 minutes, both of which were intermittently interrupted to sample responses to thought probes. We chart the evolution of cardiac dynamics over the course of the session, comparing their trajectory against behavioural measures at both the trial- and the block-level. We further analyse how behavioural performance and cardiac activity fluctuate in the 10 s preceding task-related vs. task-unrelated thought probe reports. Finally, we evaluate differences in stimulus-response processing across successive heartbeats, and across different phases within the cardiac cycle.

6.2 Methods

6.2.1 Participants

The analysis reported in this manuscript is based on a subset of individuals ($n = 24$) who had their electrocardiogram (ECG) recorded while participating in a broader study of the psychophysiological correlates of attention and mind-wandering (see Andrillon et al. 2021a). All participants were right-handed, and reported no history of psychiatric disorder or substance abuse. One participant was excluded from analysis due to a poor quality ECG record. The remaining 23 participants comprised nine females and 14 males aged 23 to 38 years ($M = 30.04$, $SD = 4.15$).

This protocol was approved by the Monash University Human Research Ethics Committee (Project ID: 10994).

6.2.2 Materials and apparatus

Task instructions and stimuli were presented using the Psychophysics Toolbox (v3.0.14; [Brainard 1997](#)) for MATLAB R2018b (The MathWorks Inc., Natick, MA, USA). Participants viewed task stimuli while seated in a dimly-lit room with their head stabilised on a support positioned approximately 60 cm from the computer screen.

All participants performed two versions of a modified SART: one featuring face stimuli, and another featuring numerical digits ([Figure 6.1a](#)). Face stimuli were sampled from the Radboud Face Database ([Langner et al., 2010](#)). Eight faces (4 female) featuring a neutral expression were selected as non-target (i.e. Go) stimuli; one image of a smiling female was selected as the target (i.e. NoGo) stimulus. Stimuli for the Digit SART were computer-generated integers ranging from 1 to 9, with “3” serving as the target stimulus. All digits were presented at a uniform font size.

Stimuli were continuously presented at the centre of the screen, with successive stimulus onsets occurring at intervals of 750 to 1250 ms (randomly sampled from a uniform distribution). All 9 stimuli within the stimulus set were present once in a pseudo-randomised order, before the set was randomly permuted for the next 9 trials (with the constraint that the same stimulus never appeared twice in succession). The probability of a target stimulus was thus fixed at 11%.

The SART was intermittently interrupted by the presentation of the word “STOP” and an accompanying sound, followed by a sequence of 7 or 8 thought probes. The present study focuses on participants’ subjective report about their attentional focus “just before the interruption”. Participants responded to this probe by selecting one of the following options: (1) “task-focused” (i.e. on-task), (2) “off-task” (i.e. focused on something other than the SART), (3) “mind blanking” (i.e. focused on nothing), (4) “don’t remember”. In the analyses that follow, responses (2) to (4) were collapsed into a single “off-task” category (but see [Andrillon et al. 2021a](#), for analysis of mind-wandering and mind-blanking as distinct phenomena). Data from the final probe, which asked participants to rate their level of vigilance “over the past few trials” on a 4-point Likert scale (1=“Extremely Sleepy”; 4=“Extremely Alert”), are also briefly reported.

6.2.3 Procedure

Following consent procedures, participants were set up for multi-modal electrophysiological signal acquisition (ECG, electroencephalogram, and eye-tracking). They were then informed of the SART and thought sampling procedures. Participants were instructed to

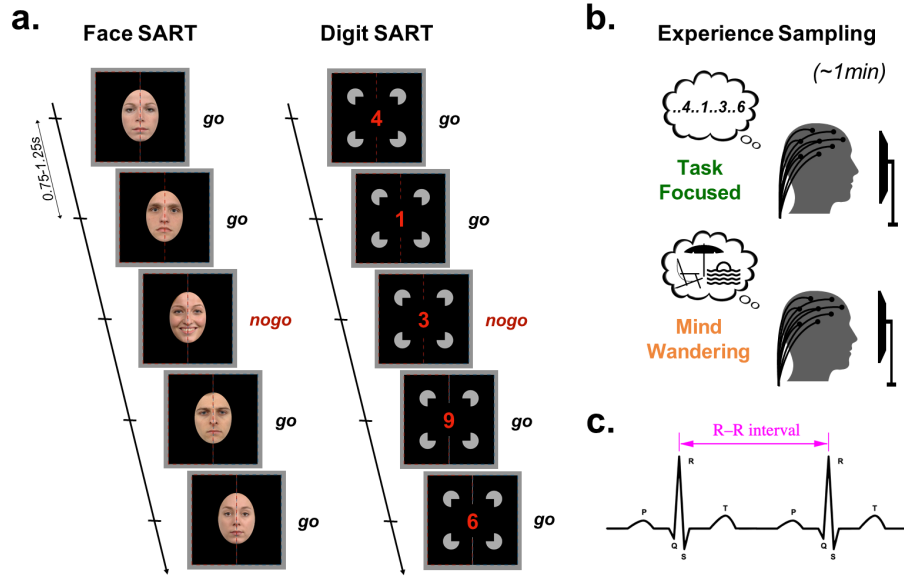


FIGURE 6.1: Schematic of experimental protocol. **a.** Participants performed two versions of a modified Sustained Attention to Response Task (SART): one comprising 9 Face stimuli (NoGo = Smile); one comprising 9 Digit stimuli (NoGo = 3). Each successive trial stimulus was presented on screen for 0.75-1.25 s. **b.** Each SART block was interrupted 10 times to undertake experience sampling. See main text for details. **c.** Schematic ECG trace depicting two successive heartbeats (R-wave peaks). The duration between each successive R-peak defines an inter-beat interval (IBI), the basic unit of analysis for the cardiac parameters investigated in this study.

pay attention to each presented stimulus and respond via button-press each time a non-target (Go) stimulus appeared. Conversely, they were told to withhold their response whenever a target (NoGo) stimulus occurred. Participants were advised to prioritise accuracy over speed of response, in line with evidence that this strategy improves the precision of SART errors as an index of attentional lapses (Seli et al., 2012).

Prior to beginning the experiment, participants performed two blocks of practice trials (1 block of 27 trials for each variant of the SART). Performance feedback (proportion of correct responses, average response time) was provided at the end of each practice block. Participants then performed 3 blocks of each SART (block order randomly permuted per participant). Blocks lasted approximately 12 to 15 minutes each and were separated by self-paced breaks (total duration: $M = 95.10$ min, $SD = 10.40$). Each block comprised 10 sets of thought probes separated by intervals ranging 30 to 70 s (randomly sampled from a uniform distribution).

6.2.4 Electrophysiological signal acquisition and preprocessing

The ECG was recorded from two electrodes placed over the deltoid muscles of each shoulder. Signals were digitized at 500 Hz and amplified by a BrainAmp system (Brain Products GmbH) as part of a high-density electroencephalography montage (BrainVision Recorder v1.21.0402; Brain Products GmbH).

Offline processing of the raw ECG was performed in MATLAB R2019b (v9.7.0.1319299) in conjunction with the EEGLAB toolbox (v2019.1; [Delorme and Makeig 2004](#)). The ECG signal was subjected to separate high-pass (passband edge = 0.5 Hz, transition width = 0.5 Hz) and low-pass (passband edge = 40 Hz, transition width = 10 Hz) filters implemented using the ‘pop_eegfiltnew’ function of the *firfilt* plugin (v2.4). The ECG record was then segmented to exclude non-task-related data.

R-wave peaks were automatically extracted from the ECG using the QRS beat-detection algorithm implemented in the ‘ConvertRawDataToRRIntervals’ function of the PhysioNet Cardiovascular Signal Toolbox (v1.0; [Vest et al. 2018](#)). The timing of resolved peaks was appended to the stimulus event information captured during online signal acquisition. Behavioural responses to SART trials and thought probes were also imported into the event structure. This information was used to map behavioural responses to stimulus and cardiac events, forming the basis of the analyses that follow.

6.2.5 Data analysis

All statistical analyses were conducted in *R* (v3.6.2; [R Core Team 2019](#)) using the *RStudio Desktop* IDE (v1.2.5033; [RStudio Team 2015](#)). Behavioural and psychophysiological data are modelled under the generalized additive mixed model (GAMM) framework.

6.2.5.1 Behavioural analysis

Behavioural performance on the SART was quantified in terms of reaction time and response accuracy.

Reaction times were calculated relative to the nearest stimulus onset preceding the registered response. Given the temporal constraints imposed by such factors as neural conduction delays, very fast responses (< 200 ms) were deemed unlikely to be representative of the generative process underpinning the rest of the response time distribution. Rather, these data are likely to comprise a mixture of anticipatory and delayed responses

from the previous trial (see, e.g., [Cheyne et al. 2009](#)). These trials were therefore excluded from all behavioural analyses, except where modelled separately as a special class of inappropriate responding.

Responses on Go trials immediately following NoGo trials were excluded from all reaction time models on the basis they might be contaminated by attentional orienting and/or post-error processing. Remaining reaction times were regressed onto block number (ordered categorical variable), stimulus type (Digit, Face), and trial type (Go, NoGo). Penalised factor smooths were included within all models to account for idiosyncratic variation of reaction times over the course of the experiment (see also [Baayen et al. 2017](#); [Cross et al. 2020](#)).

Due to the autocorrelated nature of reaction time data, first-order autoregressive (AR(1)) models were specified in order capture and adequately account for this residual structure. Note that this strategy enforced the assumption of a Gaussian (identity-linked) distribution. Since reaction time distributions were not heavily skewed (recall that task instructions emphasised accuracy over speed), and given that model diagnostics did not reveal evidence of substantial residual deviation, reaction times were not transformed prior to analysis.

Response accuracy was quantified according to the tenets of signal detection theory (SDT), a principled framework for modelling perceptual decision-making under uncertainty ([Green and Swets, 1966](#); [Macmillan and Creelman, 2005](#); [Stanislaw and Todorov, 1999](#)). Correct detections of target (NoGo) stimuli were classed as hits; failures to withhold responses to target stimuli (errors of commission) were classed as misses. Responses to non-target (Go) stimuli were classed as correct rejections; failures to respond to non-target stimuli (errors of omission) were classed as false alarms.

Sensitivity to trial type was estimated using the discriminability index d' , which was obtained by subtracting the z-scored false alarm rate from the z-scored hit rate (loglinear correction for extreme values applied in all cases; [Hautus 1995](#)). A d' of 0 indicates chance-level discrimination of targets from non-targets; more positive values are indicative of increasing sensitivity.

Response bias (decision criterion) was estimated with c , the mean of z-scored hit and false alarm rates. A c of 0 indicates no bias towards either stimulus category; more positive values indicate increasing bias towards the target category (liberal criterion), while more negative values indicate increasing bias towards the non-target category (conservative criterion).

Sensitivity and criterion measures were estimated for each participant at the block level, and regressed onto block number and stimulus type.

6.2.5.2 Thought probes

The prevalence of on-task vs. off-task attentional states over the course of the task was modelled on a probe-by-probe basis. A Binomial (logit-linked) model was used to regress each probe response as a smooth function over probe number (summing probes across blocks). Factor smooths were also applied across individual responses on successive probes. A similar analysis was performed on alertness ratings, with the exception that a distribution family suitable for ordered categorical data was selected in place of the Binomial family.

To assess differences in behavioural performance as a function of attentional state, behavioural responses falling within the 10 s period preceding each thought probe were epoched and categorised according to the corresponding probe report. Before emulating the behavioural analyses performed on the whole-task data, the prevalence of very fast responses during on- vs. off-task epochs was analysed as a function of time-on-task. A Poisson (log-linked) model was used to estimate changes in the rate of very fast responses as a smooth function over successive trial blocks. Since the ratio of on- vs. off-task reports was expected to be non-stationary over the course of the session (given the expected increase in rate of mind-wandering over time; [Antrobus 1968](#); [McVay and Kane 2009](#); [Smallwood et al. 2004a](#); [Teasdale et al. 1993](#)), a parameter encoding the proportion of off-task reports per block was included as a covariate.

The analysis of valid reaction times captured within pre-probe epochs essentially replicated the model described above, with the additional introduction of a factor encoding attentional state. Due to the limited number of events captured within each pre-probe epoch, response accuracy was quantified in terms of trial outcome rather than sensitivity and criterion estimates. Accordingly, separate Binomial (logit-linked) models were fit to responses on NoGo trials (i.e. Hits vs. Misses) and Go trials (i.e. Correct Rejections vs. False Alarms).

6.2.5.3 Cardiac parameters

Inter-beat intervals (IBIs) were calculated as the difference between successive R peak latencies resolved by the automated detection algorithm ([Figure 6.1c](#)). IBI estimates < 300 ms or > 2000 ms were excluded as improbably short or long durations, respectively. Remaining IBIs were z-score normalised on the subject-level, and estimates $> \pm 4$ normalised units excluded from analysis.

HRV was estimated in the time-domain as the standard deviation of normal-to-normal intervals (SDNN) at both the block and the epoch-level (IBIs $> \pm 4SD$ from the mean

excluded prior to calculation). These estimates were then divided by the participant’s mean IBI to render coefficients of variation (a standardised estimate of dispersion). Standardised IBI and SDNN estimates are denoted IBI_z and IBI_{cv} , respectively.

Cardiac parameters were subjected to a similar set of analyses as the behavioural data in order to assess the communality between cardiac dynamics and overt behaviour during the SART. Block-level estimates of IBI_z and IBI_{cv} were regressed onto block number and stimulus type to establish their sensitivity to time-on-task and stimulus properties, respectively. These estimates were then introduced into the reaction time model in order to assess whether they account for additional variance in behavioural performance over and above task factors. Finally, epoch-level IBI estimates were regressed onto probe response (On vs. Off) to evaluate whether they systematically varied as a function of attentional state.

6.2.5.4 Event-related inter-beat interval analysis

Next, beat-by-beat fluctuations in IBI_z were analysed for epochs consisting of 5 consecutive IBIs, centred on the IBI in which a NoGo stimulus onset occurred. This analysis enabled us to quantify the transient cardiac deceleration typically evoked by infrequent target stimuli. Moreover, by including an interaction term encoding the difference in serial IBI duration as a function of trial outcome (Hit vs. Miss), we were able to compare the magnitude of cardiac deceleration in the context of commission error vs. response inhibition. This analysis was then repeated on the epoch-level in order to investigate whether these cardiac indices of attentional orienting and error-processing differ as a function of task engagement.

6.2.5.5 Cardiac cycle phase analysis

The final set of analyses assess the distribution of stimulus onsets and behavioural responses across the cardiac cycle. This was achieved by calculating the timing of stimulus/response onsets relative to the preceding R peak, and normalising this event latency according to the duration of the IBI:

$$latency = (T_{event} - T_{R_0}) / (T_{R_1} - T_{R_0}),$$

where T is time, R_0 is the preceding R peak, and R_1 is the subsequent R peak.

Normalised latencies were binned across 40 consecutive intervals (each bin corresponding to 9 degrees on a unit circle, or 21.5 ms for the grand average IBI of 860 ms) and converted

to count data. Poisson (log-linked) models regressed event counts onto phase bins by means of a cyclic cubic smoothing spline. The specification of this variety of smoothing spline ensured that the estimated change in event count is continuous across the first and final phase bins, as would be expected given the cyclical nature of the heartbeat (i.e. as if cardiac phase had not been ‘unwrapped’ for the purposes of analysis).

6.2.5.6 Model estimation, evaluation, and visualisation

All models were estimated with the ‘bam’ function of the *mgcv* package (v1.8-33; Wood 2011). Models were fit with random effects smooths including random slopes for parametric terms, and by-participant random intercepts or factor smooths. Unless otherwise specified, smoothing terms were fit with penalised low-rank thin-plate regression splines (Wood, 2003, 2017).

Functions from the *itsadug* package (v2.3; van Rij et al. 2017) were used to fit and evaluate AR(1) models. The ‘compareML’ function from this package was used to perform model selection on the basis of fREML score. Model criticism was aided by helper functions from the *mgcViz* package (v0.1.6; Fasiolo et al. 2018).

All reported parametric effects are estimated marginal means obtained (and statistically evaluated) via the *emmeans* package (v1.5.1; Lenth 2020). Model visualisation was aided by the *tidyverse* (v1.3.0; Wickham et al. 2019), *ggeffects* (v0.16.0; Lüdtke 2018), and *cowplot* (v1.1.0; Wilke 2020) packages.

6.2.6 Data availability statement

The data analysed in this study are openly available from the OSF platform (<https://osf.io/ey3ca>; Andrillon et al. 2021b).

6.3 Results

6.3.1 Behavioural performance

In total, 1246 (1.9%) trials were rejected from analysis on account of very fast reaction time latencies. These fast responses were more prevalent during the second half of the experiment, with 58.3% of rejected trials deriving from the latter 3 SART blocks. Although the majority of rejected trials were responses to Go stimuli (78.3%), NoGo trials elicited a higher rate of rejection (3.7% vs. 1.4%). Independent of trial type, 73.7%

of rejected reaction times were responses to Face stimuli (2.5% of Face trials vs. 0.9% of Digit trials).

6.3.1.1 Reaction time

Quantile plots summarising reaction time distributions on both the individual- and the group-level are displayed in [Figure 6.2](#). These plots suggest responses to NoGo trials were faster on average than responses to Go trials (as indicated by a lower offset on the y-axis), consistent with their higher rate of rejection. Differences in the distribution of NoGo reaction time data across Face and Digit stimuli were inconsistent amongst participants, and are difficult to assess given the relatively limited occurrence of such errors. Go trial distributions, however, evinced stronger evidence of reaction time increase (i.e. slowing) in response to Face stimuli (as indicated by a higher y-axis offset and steeper slope; e.g., subplot #09).

Reaction time modelling confirmed significant main effects of task stimulus, $F(1) = 72.65, p < .001$, and trial type, $F(1) = 34.18, p < .001$, as well as a significant stimulus \times trial type interaction, $F(1) = 150.17, p < .001$. As suggested by the quantile plots, reaction times were slower on Go trials than NoGo trials, especially in response to Face stimuli (see [Figure 6.3](#), left panel). Responses on NoGo trials featuring Face stimuli were slower than those featuring Digit stimuli, $t(57534) = 2.92, p = .018$, but not significantly different from Go trials involving Digit stimuli $t(57534) = 1.44, p = .478$. It is also notable that Go trials in the Digit condition elicited markedly slower reaction times on average than those typically reported for the standard SART (e.g., 375 ms; [Manly et al. 2000](#)), consistent with the instruction to prioritise accuracy over speed ([Seli et al., 2012](#)).

The reaction time model also revealed a significant time-on-task effect, $F(5) = 4.72, p < .001$, and the modulation of this effect by trial type, $F(5) = 9.88, p < .001$. These effects are visualised in [Figure 6.3](#) (right panel). Go trial responses were consistently slower than NoGo responses over the course of the session, but tended to speed up over the second half of the task. NoGo response times, by contrast, showed more of an inverted-U pattern; NoGo errors were fastest in the first block, but slowed markedly over the subsequent two blocks. NoGo responses showed a similar rate of acceleration over the second half of the task as Go trials.

6.3.1.2 Response accuracy

Response accuracy on target (NoGo) and non-target (Go) trials is captured by hit and false alarm rate, respectively. Hit rate ranged from 46.5 to 91.1% ($M = 70.5\%$, $SD =$

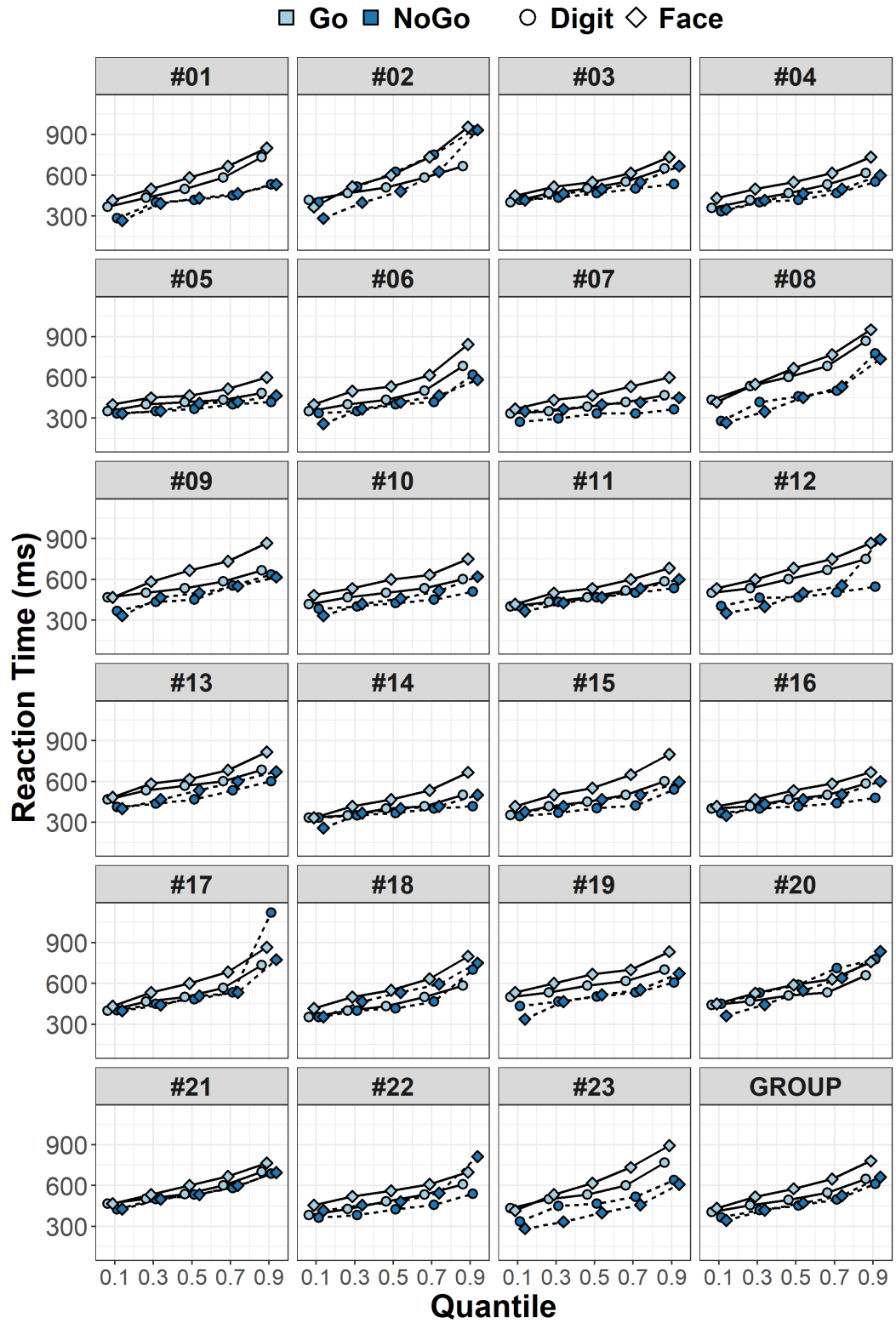


FIGURE 6.2: Individual- and group-level quantile plots depicting average reaction time distributions on the SART, factorised by trial and stimulus type. Go trials are indicated by sky blue symbols/solid lines; NoGo trials are indicated by dark blue symbols/broken lines; Digit stimuli are indicated by circle-shaped symbols; Face stimuli are indicated by diamond-shaped symbols.

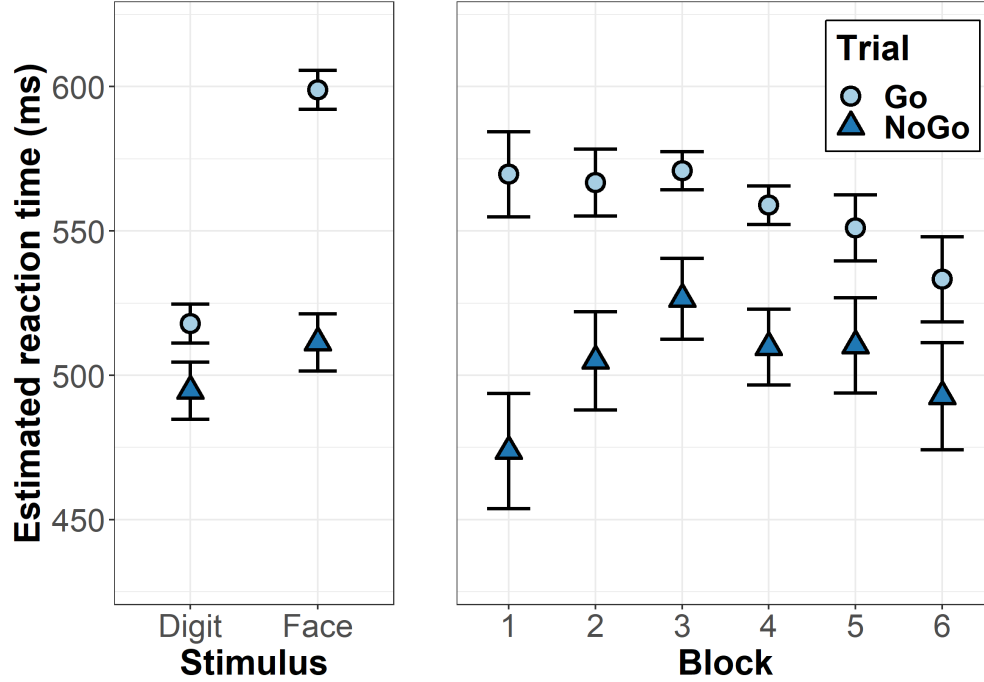


FIGURE 6.3: Estimated marginal mean reaction time (ms) on Go trials (sky blue circles) and NoGo trials (dark blue triangles) as a function of stimulus type (left panel) and time-on-task (right panel). Error bars indicate 95% confidence intervals.

10.7) across participants, while false alarm rate ranged from 1.1 to 7.5% ($M = 2.5\%$, $SD = 1.5$). Although hit rates were similar across stimulus conditions (Digit: $M = 70.5\%$, $SD = 11.7$; Face: $M = 70.5\%$, $SD = 11.3$), false alarm rate tended to be higher and more variable during the Face version of the SART (Digit: $M = 1.8\%$, $SD = 1.2$; Face: $M = 3.2\%$, $SD = 2.2$).

Signal detection theoretic measures of response accuracy (which combine information about hit and false alarm rates) were indicative of performance differences across stimulus conditions. Individual-level estimates displayed in Figure 6.4 suggest targets were easier to discriminate from non-targets in the Digit SART, although the broader dispersal of the associated marginal density is suggestive of greater inter-individual variability. Uniformly negative response criterion estimates imply that all participants adopted a conservative decision threshold, as commonly observed for infrequent target stimuli. Marginal densities indicate that participants' response criterion shifted more negatively during the Digit stimulus condition than the Face condition.

Mixed-effects models for sensitivity and criterion estimates are visualised in Figure 6.5. The sensitivity model confirmed a significant effect of stimulus condition, $F(1) = 9.88$, $p = .002$. The switch from Face to Digit stimuli improved the discrimination of targets from non-targets by an average 0.24 z-units, $t(103) = 4.13$, $p < .001$. This model also revealed

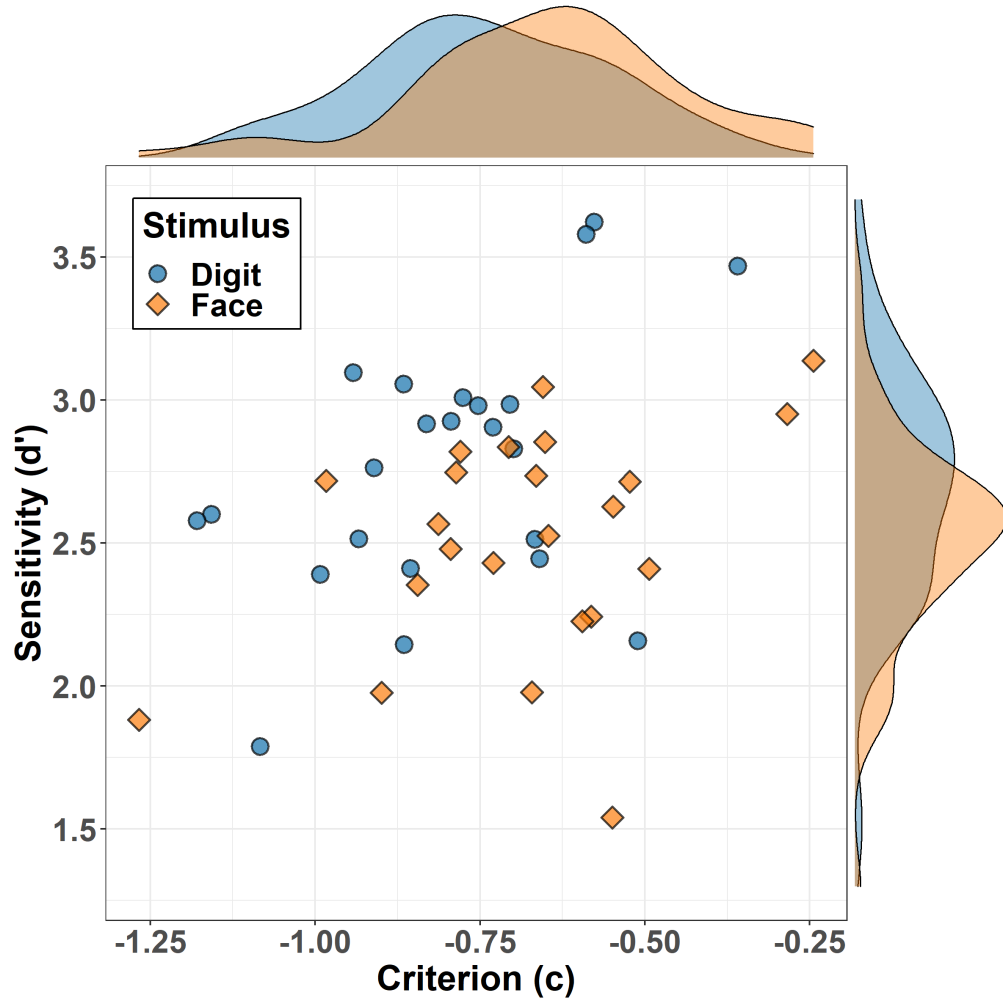


FIGURE 6.4: Individual-level estimates of response bias/criterion (c) and sensitivity (d') measures of response accuracy on Digit (blue circles) and Face (orange diamonds) versions of the SART.

a significant time-on-task effect, $F(5) = 2.63, p = .028$. Sensitivity was at its peak in the first block of trials, and decreased linearly thereafter.

The criterion model likewise revealed a significant effect of stimulus condition, $F(1) = 14.36, p < .001$, and time-on-task, $F(5) = 2.87, p = .018$. Switching from Face to Digit stimuli resulted in an average criterion shift of -0.13 z-units, $t(101) = 5.24, p < .001$. Similar to sensitivity, criterion declined linearly over the course of the experimental session.

6.3.2 Attentional states

Across 60 thought probes, participants reported being off-task ('Off') as few as 0 and as many as 55 times ($Mdn. = 34, IQR = 19$). Both the prevalence and individual variability of off-task reports are consistent with previous studies (e.g., Kane et al. 2007;

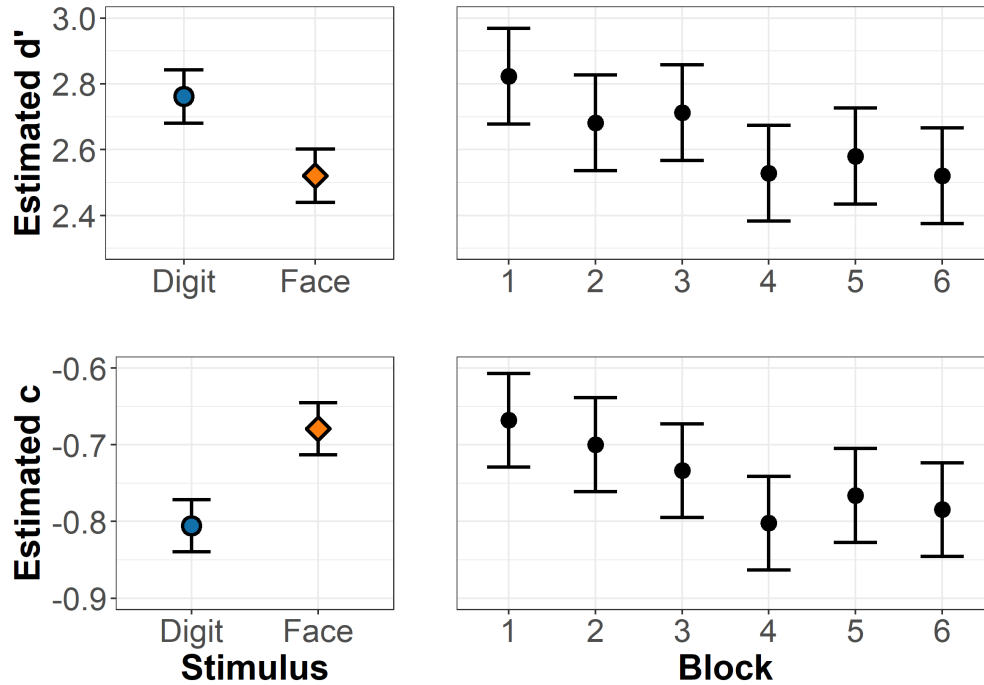


FIGURE 6.5: Estimated marginal mean sensitivity (d' ; top row) and criterion (c ; bottom row) as a function of stimulus type (left column) and time-on-task (right column). Error bars indicate 95% confidence intervals.

Killingsworth and Gilbert 2010). The distribution of on- vs. off-task reports was similar across stimulus conditions (Digit-Off: $Mdn.$ = 16, IQR = 10; Face-Off: $Mdn.$ = 17, IQR = 8).

Modelling attentional state reports on a probe-by-probe basis revealed a significant non-linear effect of time-on-task, $\chi^2(3.93) = 36.06, p < .001$. The expected probability of an off-task report increased from 37% at the end of block 1 to 60% at the end of block 4, plateaued over block 5, and declined slightly over the final block (Figure 6.6). Allowing estimates to vary by stimulus type did not significantly affect this pattern (nor indeed improve model fit).

The increased incidence of off-task reports over the course of the task was broadly in agreement with participants' subjective alertness ratings. These reports revealed a rapid decline in subjective alertness over the first two trial blocks, followed by a more gradual decline thereafter, $\chi^2(4.77) = 44.26, p < .001$.

6.3.2.1 Behavioural performance during probe-defined epochs

This section briefly rehearses the behavioural analysis reported above on the 10 s epochs preceding each thought probe. This subset of SART data comprised a total 11893 Go trials and 1472 NoGo trials, which amounted to a median 517 (IQR = 11) Go and 65

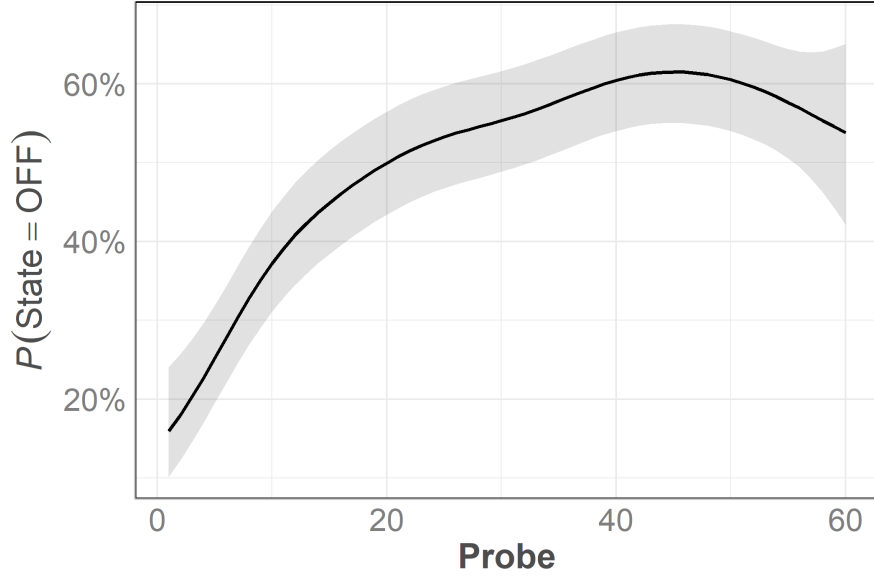


FIGURE 6.6: Probability of reporting an off-task (as opposed to on-task) attentional state as a function of time-on-task (defined here in terms of probe number; 10 probes per trial block). Shading indicates 95% confidence intervals.

($IQR = 5$) NoGo trials per participant. Of the total 11590 responses elicited over these trials, 274 (2.4%) occurred faster than the 200 ms rejection threshold applied above ($Mdn. = 10$, $IQR = 6$). The marginal distributions of these trials across response and stimulus conditions were similar to those reported in the whole-task analysis (Go: 1.7%, NoGo: 4.6%; Digit: 1.1%, Face: 3.0%).

Consistent with the tendency for reaction times to speed up as a function of time-on-task, rate of very fast responding increased linearly over time. Since the distribution of on- vs. off-task reports varied as a function of time-on-task (i.e. off-task reports became more prevalent over time), the interaction between attentional state and time-on-task on the frequency of very fast responses was modelled while holding the rate of off-task reports constant. This model revealed that very fast responses occurred with approximately equal frequency across epoch types during the first block, but occurred more frequently over time during off-task epochs, $\chi^2(1.23) = 13.73, p < .001$ (Figure 6.7). By contrast, rate of very fast responding remained approximately constant as a function of time across epochs preceding an on-task probe report, $\chi^2(1) = 0.83, p = .359$.

The reaction time model reported for the whole-task analysis was re-fit to the epoched data, broadly replicating the trial type \times stimulus condition and trial type \times time-on-task interactions described above. Although the introduction of attentional state as a fixed effect significantly improve model fit, $\chi^2(13) = 46.15, p < .001$, neither main effect nor interaction terms involving this variable significantly predicted reaction time.

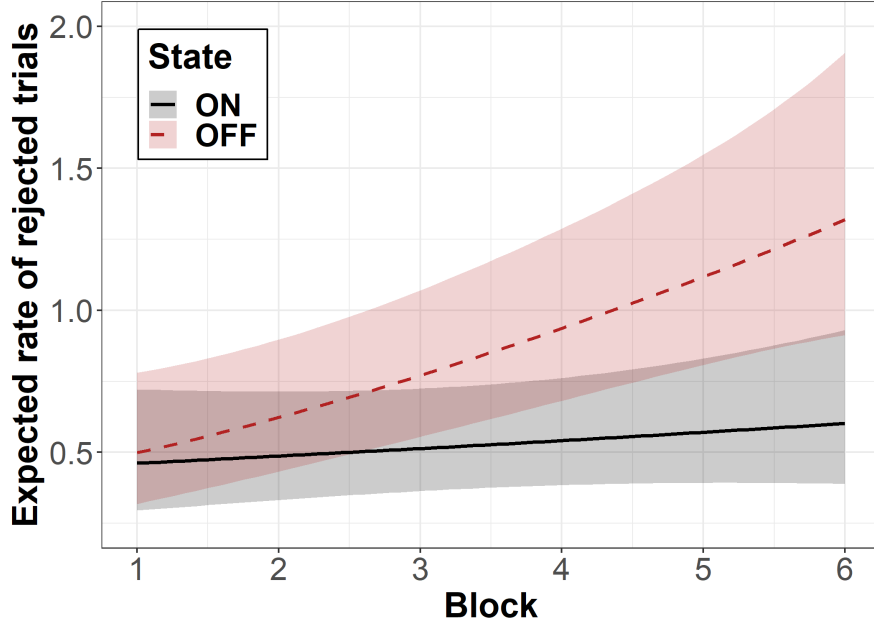


FIGURE 6.7: Expected rate of rejected trials (reaction time < 200 ms) as a function of time-on-task (smoothed across blocks) for epochs preceding on- (solid black line) and off-task (broken red line) attentional state reports. Shading indicates 95% confidence intervals.

The NoGo model revealed a significant effect of time-on-task, $\chi^2(1) = 9.42, p = .002$, whereby the probability of a Miss increased linearly over the session (consistent with the decrease in both d' and c over blocks). This model also revealed an independent effect of attentional state, $\chi^2(1) = 16.06, p < .001$, whereby the probability of a Miss increased on average from 19% during on-task epochs to 34% during off-task epochs. Neither of time-on-task nor attentional state significantly predicted Go trial performance (although note that the extreme disparity between the number of Correct Rejections and False Alarms may have limited the sensitivity of this analysis).

6.3.3 Cardiac parameters

Across individuals, mean IBI during the SART ranged from 632 to 1162 ms ($M = 860$ ms, $SD = 147$; $\sim 52 - 95$ bpm); HRV (SDNN) ranged from 26 to 126 units ($M = 63.3, SD = 25.9$).

Mixed-effects models revealed significant time-on-task effects for block-level estimates of mean IBI_z , $F(5) = 7.15, p < .001$, and IBI_{cv} , $F(5) = 4.94, p < .001$. As depicted in Figure 6.8, IBIs tended to lengthen and become more variable over the course of the SART. Notably, the tendency of mean IBI_z to increase over the first 5 blocks, before declining slightly in the final block, mimicked time-evolving changes in the probability

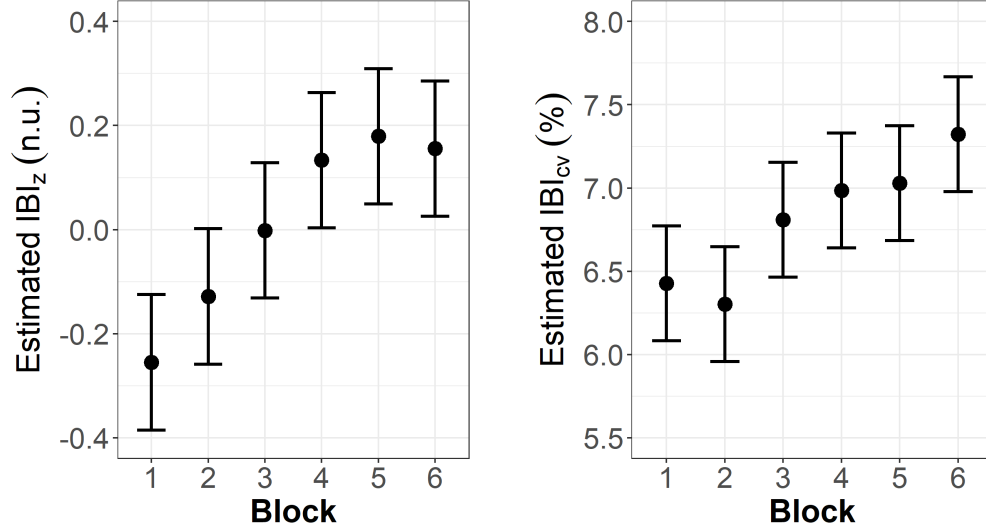


FIGURE 6.8: Estimated marginal mean inter-beat interval (left panel; z-normalised) and variability (right panel; coefficient of variation) as a function of time-on-task. Error bars indicate 95% confidence intervals.

of an off-task report (Figure 6.6). The addition of stimulus type failed to significantly improve the fit of either model.

6.3.3.1 Cardiac parameters and behavioural performance

Block-level estimates of IBI_z and IBI_{cv} were introduced into the reaction time model by means of a tensor product smoother. This additional term resulted in a significant improvement in model fit, $\chi^2(5) = 20.12, p < .001$. The interaction effect encoded by this smoother indicated that reaction time tended to decrease when IBIs were both longer and less variable, $F(4.64) = 2.41, p = .028$. This effect is consistent with the notion that sensorimotor integration involving attentive observation and rapid responding is accompanied by slower, less variable heartbeat dynamics.

Neither the sensitivity nor the criterion model were significantly improved by the inclusion of IBI_z or IBI_{cv} (jointly or independent of one another).

6.3.3.2 Cardiac parameters and attentional state

Turning next to the attentional state reports, mean IBI_z and IBI_{cv} were calculated over the 10 s preceding each probe and regressed onto variables encoding time-on-task and attentional state. Both of these short-term cardiac estimates showed very similar patterns of temporal fluctuation as compared to the whole-task analysis, irrespective of whether time-on-task was modelled at the probe- or the block-level. However, neither

model indicated significant effects of attentional state, suggesting that normalised IBI duration and variation did not significantly differ across on- vs. off-task epochs. These results remained essentially unchanged after models were re-fit using estimates calculated (1) from the final five IBIs preceding probe onset (on the assumption these IBIs may have provided a more sensitive reflection of the probe report), and (2) after the exclusion of IBIs containing or within two beats following a NoGo stimulus onset (on the assumption that orienting- and/or error-related processing may have diluted attentional state differences).

6.3.4 Event-related inter-beat interval analysis

The next set of analyses investigate beat-by-beat changes in IBI_z that occurred within ± 2 intervals of NoGo trial onset. IBI_z was regressed onto an ordered factor indexing the serial position of successive IBIs relative to the IBI in which a NoGo stimulus onset occurred (indexed as IBI_z^0).

IBI_z varied significantly as a function of IBI sequence, $F(4) = 131.42, p < .001$, and was further modulated according to trial outcome (i.e. Hit vs. Miss), $F(4) = 4.52, p = .001$. This interaction effect is visualised in [Figure 6.9](#). As expected, this figure indicates that the IBI immediately following a NoGo trial (IBI_z^{+1}) is markedly longer than those preceding it (IBI_z^{-2}, IBI_z^{-1}), and indeed, than the IBI in which stimulus onset occurred (IBI_z^0).

Post-hoc contrasts revealed no significant differences between IBI_z^{-2}, IBI_z^{-1} , and IBI_z^0 , irrespective of NoGo trial outcome. Although IBIs prior to stimuli that evoked Misses (commission errors) were numerically shorter than those preceding Hits (withheld responses), pairwise comparisons between corresponding serial IBIs did not significantly differ. However, the duration of IBI_z^0 was significantly longer when a response was successfully inhibited as opposed to when an erroneous response was elicited, $t(38126) = 3.56, p = .014$.

The two IBIs following NoGo stimulus onset were both significantly longer than the IBI in which stimulus onset occurred, $ts(38126) > 6.37, ps < .001$. While IBI_z^{+1} was also significantly longer than IBI_z^{+2} in the context of successful response inhibition, $t(38126) = 10.76$, this difference was not significant following commission error, $t(38126) = 2.52, p = .258$. This observation suggests that the marked cardiac deceleration associated with the occurrence of a NoGo trial was prolonged in the wake of erroneous responding, consistent with orienting accounts of post-error processing ([Notebaert et al., 2009](#); [Wessel, 2018](#)).

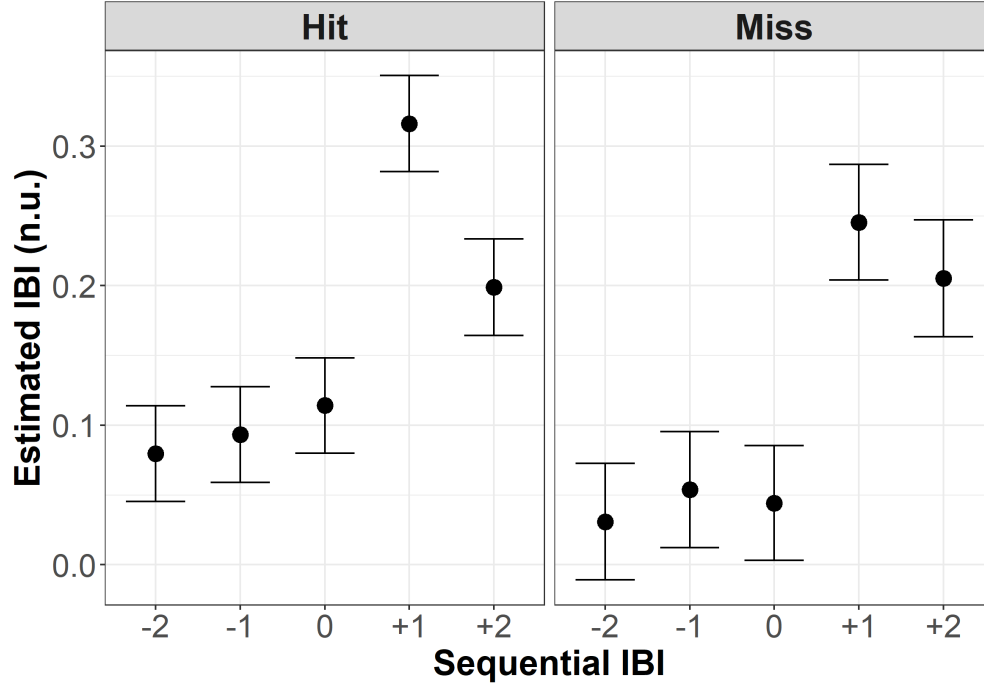


FIGURE 6.9: Estimated marginal mean inter-beat interval (IBI; z-normalised) as a function of the sequential ordering of IBIs relative to NoGo trial onset (0 = interval in which NoGo stimulus onset occurred). Error bars indicate 95% confidence intervals.

To establish whether the magnitude of the cardiac deceleration evoked by NoGo stimuli varied as a function of trial outcome, a comparison of sequential contrasts was performed. This interaction contrast revealed no significant difference in the change from IBI_z^0 to IBI_z^{+1} across Hits and Misses, $t(38126) = 1.61, p = 1$. However, the difference between IBI_z^{+1} and IBI_z^{+2} varied significantly across Hits and Misses, $t(38126) = 3.99, p < .001$, in line with the inference that post-error processing causes additional cardiac slowing over and above that evoked by the presentation of a salient stimulus.

Refitting this model to the subset of NoGo trials that fell within the pre-probe time-window resulted in a qualitatively similar profile of heartbeat fluctuation over the IBI series as in the whole-task analysis; however, only the main effect of IBI index was statistically significant, $F(4) = 17.27, p < .001$. The introduction of a variable encoding attentional state failed to significantly improve model fit, irrespective of whether this variable was allowed to interact with other fixed effects.

6.3.5 Cardiac cycle phase analysis

The event-related analysis revealed that the duration of the IBI in which NoGo stimuli were presented was longer in trials that were not responded to, as opposed to those that elicited an error of commission. One possible explanation for this observation is that

this momentary slowing of the cardiac cycle indexes (and potentially contributes to) a spontaneous phasic increase in attentional focus, thus reducing the probability that an inappropriate response will be issued. However, this interpretation is complicated by the fact that beat-by-beat cardiac dynamics are exquisitely sensitive to the timing of unfolding sensory events: stimuli occurring early in the cardiac cycle have been found to prolong IBI duration, whereas events occurring at later phases of the cycle tend to affect the subsequent IBI (and may even curtail the current IBI; [Coles and Strayer 1985](#); [Lacey and Lacey 1977, 1980](#); [Jennings and Wood 1977](#)).

To investigate whether the difference in IBI_z^0 observed across hits and misses was an artefact of chance fluctuations in stimulus onset timing, the distribution of stimulus onsets across the cardiac cycle was modelled for each subset of trials. Stimulus onset latencies relative to the preceding R-peak were binned into 40 equidistant intervals that spanned the normalised length of the IBI. The number of stimulus onsets was then estimated as a function of cardiac phase by regressing expected counts onto phase bins using a cyclic cubic smoothing spline that was allowed to vary by trial outcome (Hit vs. Miss).

The parametric term encoding the difference in stimulus onset counts between response outcomes was significant, $\chi^2(1) = 99.04, p < .001$, reflecting the fact that NoGo trials yielded hits more frequently than misses. Crucially, however, the model revealed no evidence of non-uniformity in the distribution of stimulus onsets across phase bins, irrespective of whether responses to stimuli were correctly withheld or not. It thus seems unlikely that the longer duration of IBI_z^0 observed for NoGo trials that did not elicit a response was the result of chance differences in the distribution of stimulus onsets across the cardiac cycle.

Intriguingly, repeating this analysis with Go stimuli (Correct Rejections only, given the scarcity of False Alarms) produced evidence that stimulus onsets were non-uniformly distributed across the cardiac cycle, $\chi^2(2.09) = 8.92, p = .005$. This was surprising, since one might expect the frequent occurrence and non-target status of these stimuli, coupled with the random jittering of successive trials, to produce an approximately uniform array of onset latencies.

To test the robustness of this finding, a set of ‘null’ cyclic cubic smoothing functions was constructed via Monte Carlo simulation. These spline functions were generated by fitting 1000 identically-specified models to pseudo-count data derived by randomly sampling a uniform distribution $i \times j$ times, where j corresponds to the number of events in the real data for each i^{th} participant). As depicted in [Figure 6.10](#) (left panel), the smooth fit to the real data exceeded the bounds constructed by taking the maximum absolute deviation of the null smooth at the 95th percentile (corresponding to $\alpha = .05$).

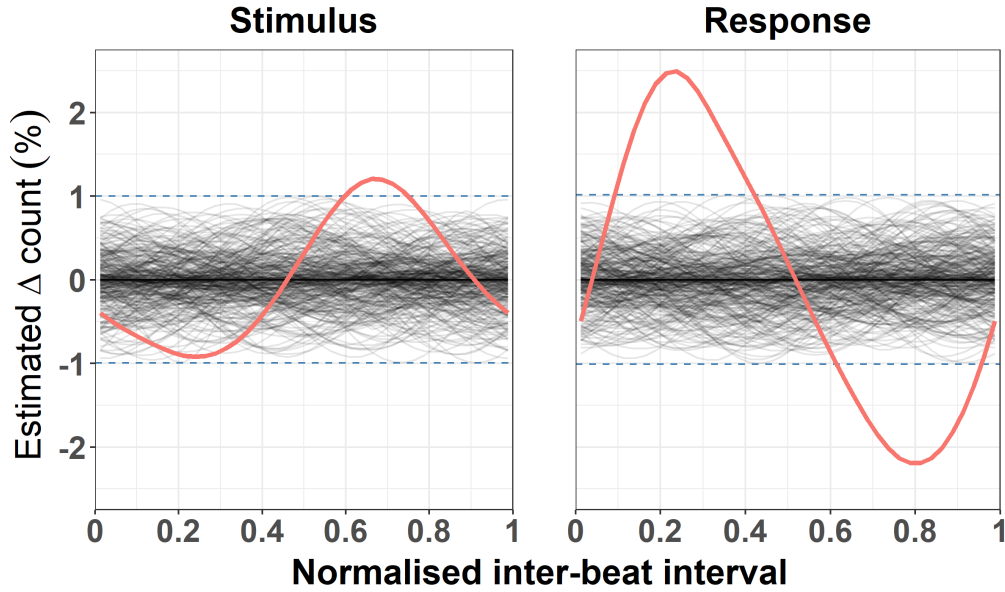


FIGURE 6.10: Estimated change in stimulus (left panel) and response (right panel) onset counts as a function of normalised cardiac cycle phase. Red function estimated from empirical counts, grey functions estimated from Monte Carlo simulated counts. Broken lines indicate maximum deviation of simulated functions at the 95th percentile.

This finding suggests that the event latency data modelled above were unlikely to have been generated by a uniform process.

Next, the cardiac cycle phase analysis performed on Go trial stimuli was repeated for participant responses. The cyclic cubic smoothing spline was again significant, $\chi^2(2.77) = 19.14, p < .001$, indicative of a non-uniform distribution of response onsets across the cardiac cycle. Whereas the stimulus onset analysis revealed a tendency for stimuli to cluster about the later (i.e. diastolic) phase of the cycle, responses were more frequently issued in the early (i.e. systolic) phase of the cycle. Monte Carlo simulation-based analysis again confirmed that the magnitude of the estimated deviation from uniformity was unlikely to have occurred by chance (see Figure 6.10; right panel). No such effects were detected for responses on NoGo trials, nor for responses that were faster than 200 ms; however, both of these supplementary analyses suffer from substantial reductions in statistical power.

Finally, the cardiac cycle models reported above were refit to pre-probe data to investigate whether the phase-dependent patterning of stimulus and response onsets across the cycle was modulated by attentional state. These models also furnished evidence of non-uniformly distributed latencies; however, this effect only manifested after smoothers were allowed to vary by attentional state. Perhaps somewhat counter-intuitively, stimulus and response onset latencies were both uniformly distributed across the cardiac cycle in the lead-up to an on-task probe response, while off-task epochs rendered non-linear

estimates qualitatively consistent with those reported above. This effect was not statistically significant in the case of the stimulus onset model, $\chi^2(1.66) = 3.90, p = .073$, but was significant in the case of the response onset model, $\chi^2(1.63) = 4.40, p = .047$.

6.4 Discussion

This study set out to examine the cardiac correlates of attentional fluctuation across multiple timescales, focusing in particular on the contrast between episodes of task-unrelated vs. task-focused cognition. Cardiac measures derived from electrocardiograms collected during the performance of a Sustained Attention to Response Task (SART) revealed a tendency for the cardiac rhythm to become slower and more variable over the course of the task. While these measures were not sensitive to the modulation of behavioural performance by differing stimulus types, they jointly accounted for unique variance in trial-level reaction time data.

Although block-level changes in cardiac pacing over the course of the experimental session approximated time-evolving changes in the propensity for task-unrelated thought, short-term indices of cardiac dynamics revealed little evidence of systematic variation with self-reported attentional states. One interesting exception to this pattern was identified on the level of the cardiac cycle: Whereas stimulus and response events were uniformly distributed across the cycle for heartbeats that occurred in the time-window preceding ‘on-task’ reports, they appeared to cluster around a particular phase of the cycle during heartbeats preceding ‘off-task’ reports. The apparent specificity of this patterning to periods of task-unrelated thought raises interesting questions about the heart’s sensitivity to sensorimotor events under different attentional regimes.

The SART further availed the opportunity to investigate fine-grained differences in cardiac activity around the onset of target (i.e. NoGo) stimuli, and in particular, cardiac fluctuations following withheld vs. erroneous responses. Two key findings from this analysis were that (1) inhibited responses to NoGo trials were associated with slower inter-beat intervals (IBIs) at trial onset, and (2) similar degrees of cardiac slowing were observed in the interval following stimulus onset, despite differing behavioural outcomes. Decomposing NoGo trials by attentional state revealed no significant difference in the magnitude of the cardiac deceleration evoked by target stimuli and/or erroneous responses between on- and off-task, although this null finding should be treated as somewhat provisional given the limited number of trials included in the analysis.

6.4.1 Time-on-task effects in the SART

As mentioned in the Introduction, the response format of the SART paradigm can be construed as an inversion of the canonical Continuous Performance Test (CPT), both of which were initially developed to detect attentional deficits in brain injured individuals (Robertson et al., 1997; Rosvold et al., 1956). One advantage of the SART over the CPT (indeed, one of the primary motivations for its development; Robertson et al. 1997) is its capacity to render relatively rich information about the fluctuation of attention over relatively short periods of time (i.e. < 5 min). Consequently, the behavioural and physiological effects of prolonged SART performance have received scant attention (cf. Bonnefond et al. 2010; Foxe et al. 2012; Staub et al. 2015).

The SART is a popular method in laboratory-based mind-wandering research, where it may be extended over longer periods than are typically necessary for the purposes of characterising attentional deficits in clinical and at-risk populations. Within this context, the SART might simply be viewed as a reliable means of eliciting episodes of task-unrelated thought; debates about the extent to which it taps attentional (as opposed to inhibitory, motoric, and ancillary) mechanisms might be of peripheral concern. Nonetheless, the behavioural data afforded by such tasks are often used in conjunction with thought sampling techniques to draw inferences about cognitive dynamics during task performance. It is therefore important to consider how performance on longer variants of the SART is modulated by time-on-task, fatigue, and/or ultradian rhythms – processes that might exert significant non-stationary or non-linear influences on attention and arousal.

Previous studies attempting to minimise the confounding of attentional and inhibitory control failures on SART performance have demonstrated that the instruction to optimise accuracy over speed reduces error rates relative to standard procedures (Seli et al., 2012). This boost in accuracy is however somewhat transitory: error rates significantly increased from the first to the second half of a ~ 14 min SART; mean reaction time was non-significantly reduced. The present findings replicate and extend this observation by revealing that the performance of a protracted SART protocol under ‘go slow’ instruction engenders a steady decline in sensitivity reminiscent of the classic ‘vigilance decrement’ (Mackworth, 1956; Parasuraman, 1979). This effect was accompanied by a concomitant decline in response criterion, suggestive of a gradual shift towards a more conservative response strategy.

Our reaction time analysis further suggests that the aforementioned criterion shift might be partially explained by increasing automatization of response behaviour over the latter half of the task. This time-on-task effect seems to run counter to consistent reports of

the temporal stability of mean reaction times to Go stimuli on the SART ([Cheyne et al., 2011](#); [McVay and Kane, 2009](#); [Seli et al., 2012](#); [Staub et al., 2015](#)); however, this apparent contradiction is resolved by remarking that these studies involved protocols that were probably too short (< 40 min) for this effect to manifest. Together with the observation that reaction times on NoGo trials slowed substantially over the first 3 blocks, before speeding up again over the remaining 3, these findings suggest some degree of time-evolving, non-stationary fluctuation in response strategy persists in spite of the instruction to optimise accuracy.

6.4.2 Cardiac and cognitive control across time

As expected, the frequency of task-unrelated thought episodes increased over the course of the session, albeit with a minor reversal of this effect in the final block (participants were aware that block 6 would be their last, which may have precipitated a boost in arousal and/or motivation). Increases in mind-wandering were accompanied by increases in the probability of inappropriate responding, as indexed by commission errors and very fast reaction times. Consistent with previous studies of cardiac activity during sustained attention (e.g., [Pattyn et al. 2008](#); [Porges 1992](#)), mean IBI and HRV also tended to increase with time-on-task – on average, cardiac dynamics become slower and more variable over time.

It is tempting to interpret these trends – and especially the similarity in the patterning of mean IBI and frequency of off-task reports over time – as indicative of a common, unitary process; however, this picture is complicated by an apparent lack of covariation amongst these variables on shorter timescales. Contrary to previous reports associating heart rate speeding (corresponding to IBI shortening) with task-unrelated thought in the SART ([Smallwood et al., 2004a](#)) and other paradigms ([Smallwood et al. 2004b, 2007](#), see also [Ottaviani et al. 2013](#)), attentional state did not significantly predict IBI duration.

Even if one were to explain away this apparent discrepancy by appealing to sampling error or methodological differences, a bigger inconsistency seemingly looms: If previous studies linked mind-wandering with increased heart rate, why should the data presented here indicate a global trend in the opposite direction?

One possibility is that the short-term effects detected in these previous studies were at least partially mediated by individual differences in mind-wandering propensity (cf. [Smallwood et al. 2004a](#)) and/or content (cf. [Ottaviani et al. 2013](#)). Another possibility is that long and repetitive paradigms such as the one employed here induce substantial adaptation effects, whereby habituation to task demands obliterate short-term state

differences that may have occurred during early phases of the task. For instance, participants might initially engage in more effortful bouts of task-focused attention after realising they had drifted off-task, precipitating a reduction in heart rate consistent with orienting and attentive observation. As the task wears on, however, depth of attentional processing during the on-task state might wane considerably, rendering its cardiac correlates indistinguishable from those of the off-task state.

This speculation might hold an important clue for explaining why cardiac activity becomes progressively slower and more variable over the course of monotonous tasks like the SART. Such vigilance tasks are inherently boring by design – simple, unengaging activities that become challenging in virtue of the tedium that must be endured in order to occasionally modulate routinised behaviour. Such conditions are of course highly conducive to mind-wandering ([Giambra, 1995](#)); stimulus processing demands are shallow and responses readily automated, while the relative absence of such collative factors as novelty and surprise (except perhaps for the occasional error of commission) render the task profoundly uninteresting. Under such circumstances, mind-wandering can be construed as a form of mental action that helps to alleviate some of the tedium of the present situation through a kind of covert ‘self-stimulation’ (cf. [Perrykkad and Hohwy 2020](#)).

In a recent study of sustained attention under perceptual uncertainty, we ([Corcoran et al., 2021](#)) suggested that short-term modulations of cardiac activity might serve to reduce internal noise in an effort to facilitate sensory processing. We propose that the long-term trends in cardiac regulation observed in the present study and elsewhere (e.g., [Pattyn et al. 2008](#)) could reflect the polar opposite of this phenomenon, whereby the heart is recruited by central processes to furnish unpredictable sensory feedback. As the individual becomes increasingly habituated to (and bored by) their task, the heart is essentially liberated to ‘explore’ its available state space in a way that is analogous to how one’s attentional focus is permitted to drift off-task and explore counterfactual states. If this interpretation is on the right track, one shouldn’t necessarily expect there to be a tight relationship between cardiac and cognitive dynamics at shorter timescales – mind-wandering and ‘heart-wandering’ may constitute essentially independent modes of covert action in response to slowly-evolving environmental dynamics.

6.4.3 Attention and behaviour in cardiac time

Setting longer-term dynamics aside, our analysis also examined how behavioural performance and heartbeat activity varied on the timescale of the IBI itself. Dealing first with fluctuations in the duration of IBIs surrounding the onset of a NoGo stimulus,

we observed a phasic lengthening of IBIs following trial onset, and the prolongation of this effect in the wake of erroneous responding. Given that the magnitude of such effects have previously been linked to error-awareness, we had hoped to examine whether they were modulated by attentional state. However, since we were unable to replicate the response-contingent differences in IBI duration in the data sampled from pre-probe epochs, it seems unlikely this analysis was sufficiently powered to detect meaningful differences across on- vs. off-task reports.

From a more methodological perspective, these findings present independent support for the claim that the ‘go slow’ version of the SART mitigates the conflation of attentional lapses with inhibitory failures from speeded responding. Previous psychophysiological research investigating the cardiac correlates of inhibitory control in the context of Go/NoGo and related paradigms has demonstrated that successful response inhibition is accompanied by more pronounced cardiac deceleration in the IBI following stimulus presentation as compared to trials in which an error of commission was made ([Jennings and van der Molen, 2002](#); [van Boxtel et al., 2001](#)). The present findings provide no such evidence of response-contingent modulation; the change in IBI duration from the interval in which the NoGo stimulus was presented to the subsequent interval did not significantly vary across Hits and Misses. This observation lends weight to the notion that SART errors are more likely to derive from attentional rather than inhibitory control failures when performed under ‘go slow’ instructions.

Conversely, our event-related analysis revealed evidence of a more subtle (but significant) difference in the length of IBIs in which NoGo stimuli were presented. As this effect did not appear to be driven by chance differences in the timing of stimulus onsets, it seems more likely that transient fluctuations in cardiac dynamics at the time of stimulus presentation index transient fluctuations in perceptual sensitivity and attentional engagement, thereby predicting trial outcome. Such phasic dynamics could emerge as a downstream consequence of central oscillatory processes governing state-level arousal or alertness (e.g., neuromodulatory gain control mechanisms), or indeed could play an active role in modulating or gating sensory input, as suggested by classical work in the psychophysiology of attention ([Lacey and Lacey, 1974, 1978](#)) and recent work in the computational neuroscience of mind-body integration ([Allen et al., 2019](#)).

If cardiac states are indeed actively modulated for the purposes of optimising sensorimotor integration ([Lacey, 1972](#); [Lacey and Lacey, 1974](#)), this might imply that some of these processes ought to be preferentially scheduled to occur in particular phases of the cardiac cycle. Such phase-dependent cycle-timing effects have garnered significant interest in recent years, especially in domains of sensory and affective processing. Although these phenomena remain poorly understood, a general picture is emerging in

which sensory processing is sharpened in the quiescent period of the cardiac cycle (diastole), and less sensitive during its active period (systole). Recent work has further indicated that voluntary actions may be preferentially executed during systole, thereby availing sensory systems with novel input at a time when they are optimally disposed to process it (i.e. in the post-action diastolic phase; [Galvez-Pol et al. 2020](#); [Kunzendorf et al. 2019](#); cf. [Herman and Tsakiris 2020](#); [Park et al. 2020](#)).

Our findings broadly conform to this pattern, with stimulus onsets preferentially occurring later in the cardiac cycle, and responses occurring earlier. The stimulus onset dependency is puzzling, however, since (contrary to response onsets) participants were unable to control the timing of these events; moreover, participants should not have been able to anticipate the precise timing of stimulus onsets, given the random jittering of the inter-stimulus interval (ISI) and the absence of any cue besides the passage of time.

One possibility is that participants anticipated the approximate timing of stimulus onset over a sufficient number of trials for the effect to emerge (for instance, by extracting the average ISI duration and entraining to this rate). Such predictions might have been augmented by the prolongation of IBIs (i.e. postponement of the next heartbeat) in anticipation of an imminent stimulus onset in longer-than-average ISIs. It's not immediately clear, however, why these adaptive effects should only manifest during periods of task-unrelated thought – indeed, it seems intuitive that being on-task should promote stronger phase-locking, given the increased attentional focus allocated to task stimuli.

One tentative explanation for why off-task states might cause the cardiac rhythm to be more susceptible to synchronisation with rhythmic environmental stimuli is that these states entail the decoupling of thought and executive control processes not only from one's external environment, but also from one's *internal* environment. Such decoupling would imply a certain ceding or devolution of control over efferent outflows to lower-level sensorimotor and autonomic networks. Thus, in much the same way as certain mind-wandering states are characterised in terms of absent-mindedness, 'running on autopilot', and the automatising of behaviour, cardiac states may become increasingly liberated from the influence of fronto-cortical regulation. Under such conditions, concurrent cycles of cardiac and somatomotor activity may be more readily integrated in relation to unfolding sensory events as they are rapidly processed and integrated by limbic and brainstem circuitry (cf. [Adelhöfer et al. 2020](#); [Larra et al. 2020](#)).

Intriguingly, then, our findings suggest that the cardiac rhythm may become increasingly coupled with environmental dynamics as higher-order brain networks move the mind further away from them. One might envisage an adaptive function that underwrites

such phenomena: As the mind decouples and disengages from its surroundings, the body becomes increasingly attuned to them, to the extent that increased visceral sensitivity to external events may serve as an ‘early warning system’ that alerts higher brain centres when something is amiss, thereby causing the reorientation of attention back towards the present situation. This kind of account fits nicely with recent work demonstrating the involvement of cardiac afferent feedback in error-processing ([Lukowska et al., 2018](#)), where the heart’s sensitivity and rapid adjustment to unexpected events might constitute sensory evidence that an error has occurred ([Hajcak et al., 2003](#); [Wessel et al., 2011](#)).

6.5 Conclusion

This study provides the first in-depth analysis of the cardiac dynamics of mind-wandering. Inter-beat intervals became longer and more variable over the course of the experiment as the incidence of mind-wandering increased; however, short-term estimates of cardiac activity did not systematically covary with attentional state. Exploratory analysis revealed evidence of a cardiac cycle-timing effect that was confined to periods of task-unrelated thought. This finding suggests previous reports of phase-dependent sensory processing and action execution may be more likely to manifest when the mind wanders.

7

Finding meaning in the noise: Expectations guide attention towards the content of degraded speech

The previous two chapters have examined the adaptation of physiological states in response to environmental uncertainty, with a particular focus on the psychophysiological relation between perception, overt behaviour, and covert action. The main message of Chapter 5 was that the brain may modulate the cardiac rhythm in order to reduce uncertainty when confronted with perceptual ambiguity. Chapter 6 added to this picture by suggesting that the brain may seek to expose itself to *increased* uncertainty (i.e. by reducing precision over the timing of interoceptive feedback) when starved of novelty.

Superficially, the idea that the brain pursues uncertainty under certain circumstances may seem antithetical to the active inference perspective; however, this interpretation is entirely in keeping with the notion that organisms periodically sample actions that *maximise* Bayesian surprise in order to resolve uncertainty (see, e.g., [Clark 2018](#); [Friston et al. 2015, 2017b, 2021](#); [Parr et al. 2019](#); [Schwartenbeck et al. 2019](#)), thereby maintaining the fitness of their generative model in a complex, volatile world ([Corcoran, 2019](#)). While acknowledging that allostatic regulation is fundamentally beholden to the pragmatic exigencies of biological viability (cf. [Seth 2015](#); [Seth and Tsakiris 2018](#)), these

findings suggest that cardiac states might yet afford their own (perhaps somewhat limited) opportunities for exploration or ‘epistemic foraging’.¹

A further contribution of Chapter 6 was its more explicit focus on the hierarchically-nested timescales across which cognitive, behavioural, and cardiac states unfold. Although this study did not uncover evidence of a systematic relation between cardiac activity and self-reported attentional states over epochs spanning several seconds, effects on both slower (i.e. successive blocks) and faster timescales (i.e. successive heartbeats; within the cardiac cycle) were revealed. Such findings speak to the multi-scale structure of organismic activity, whereby the fast dynamics of perception and action are modulated and contextualised by slowly evolving physiological and environmental processes (cf. Chapter 3).

In this final experimental chapter, the focus shifts from the nested oscillatory dynamics of the heart to those of the brain. While this move represents a departure from a core thematic component of the past three chapters, it enables a more detailed investigation of the psychophysiological correlates of attention *qua* covert mental action. By focusing on the problem of (degraded) speech comprehension, this chapter addresses several previously encountered themes from a new perspective, including: the rhythmic organisation of (covert) action and perception, the adaptation of physiological processes in response to complex environmental dynamics, and the imperative to resolve uncertainty over multiple temporal scales. These continuities, along with some important contextual information about the nature of language and its neural processing, are briefly elaborated below.

7.0.1 Language, hierarchy, and prediction

Spoken language constitutes an extraordinarily complex form of auditory stimulation, yet one that most people are remarkably adept at understanding (Scott, 2019). Besides the astonishing amount of fine-grained acoustic detail embedded within the speech signal – information that needs to be encoded at the sensory periphery, categorised into discrete units of sound (phonemes, syllables), and mapped onto semantic structures (words, phrases; Bornkessel-Schlesewsky et al. 2015; Hickok and Poeppel 2007) – successful speech comprehension may also require the listener to adjudicate amongst competing possible interpretations of the unfolding utterance. Under such circumstances, inferring the message that the speaker intended to convey depends on the listener’s ability to

¹There are some interesting questions here about the relation between physiological and behavioural variability: Could HRV reflect something like an interoceptive mode of curiosity-learning, one that is indulged in stress-free situations? Does the chronic reduction of HRV in disorders such as anxiety and depression reflect a fundamental aversion to novelty or uncertainty, as speculated in the discussion section of Chapter 5?

marshal various sources of extraneous information, including prior knowledge (e.g., of the speaker, context, and world) and nonverbal cues (e.g., prosodic features such as pitch and intonation, accompanying gestures; [Hagoort and van Berkum 2007](#); [Hagoort 2017](#)).

This cursory gloss on the intricacies of speech perception and recognition highlights two salient features of language that are of particular interest for the purposes of this chapter: (1) language is characteristically *hierarchical* in structure, consisting of nested elements that can be flexibly combined to form meaningful higher-order representations (e.g., words, sentences; [Rosen 1992](#)); (2) the semantic content of an utterance (or text) is frequently under-determined by its linguistic content, thus requiring the comprehender to make their own judicious inferences in order to ‘fill in the gaps’ ([Christiansen and Chater, 2016](#); [Hagoort, 2019](#)). As will be elaborated on below, this problem is exacerbated in the auditory modality, where the boundaries separating consecutive words need to be inferred in the absence of unambiguous acoustic markers.

A third important feature of speech resides in its essential *temporality*. The components of spoken language are necessarily articulated in an ordered sequence, necessitating in turn the incremental (re)construction of local and, in the case of sentence- or discourse-level processing, long-distance dependency relations. While this mode of presentation opens up opportunities for conveying additional information in the form of prosodic cues, it dramatically restricts the quantity of information availed to the sensorium at any given moment to that captured within a narrow, constantly moving window (cf. the ‘Now-or-Never’ bottleneck; [Christiansen and Chater 2016](#)).² Consequently, comprehension not only entails that the formidable task of encoding, recapitulating, and disambiguating the speech signal is accomplished, but also that it is accomplished in ‘real-time’ – fast enough to keep abreast of the ‘continual deluge of linguistic input’ ([Christiansen and Chater 2016](#)) unfolding before one’s ears.

In light of the hierarchical organisation of language, the role of induction in resolving ambiguity, and the sequential ordering of verbal utterances (along with the significant processing demands such ordering entails), it may come as no surprise that predictive mechanisms figure prominently within contemporary models of sentence comprehension (see, e.g., [Altmann and Mirković 2009](#); [Christiansen and Chater 2016](#); [Dell and Chang 2014](#); [Hagoort 2017](#); [Kleinschmidt and Jaeger 2015](#); [Levy 2008](#); [Pickering and Garrod](#)

²To be sure, a similar point could be made about reading; one can only sample a limited region of space on a page or screen at a time. Crucially, however, one typically sets one’s own pace when reading, thus controlling the flow of information (and indeed, enabling one to backtrack and repeat or skip ahead and omit portions of text at will). Speech rate, on the other hand, is ordinarily beyond the direct control of the listener.

2013).³ Indeed, as evidence of predictive dynamics has accrued across multiple levels of linguistic processing (see, e.g., Bornkessel-Schlesewsky and Schlewsky 2019; DeLong et al. 2014; Federmeier 2007; Hagoort and van Berkum 2007; Kamide 2008; Kuperberg 2016; Kutas and Federmeier 2011; Kutas et al. 2011; Saffran and Kirkham 2018; cf. Nieuwland 2019; Van Petten and Luka 2012), debate has shifted away from the question of whether language comprehension is (sometimes) facilitated by prediction, to that of whether comprehension is fundamentally predictive in nature (Huettig, 2015; Huettig and Mani, 2016).

Unfortunately, progress in this broader debate has been stymied somewhat by the diversity of ways in which the concept of prediction has been interpreted within the psycholinguistic literature (for discussion, see Bornkessel-Schlesewsky and Schlewsky 2019; DeLong et al. 2014; Kuperberg and Jaeger 2016). Amidst the conceptual maelstrom, however, predictive coding-informed accounts of language comprehension have gained substantial traction. Of particular interest for the purposes of this chapter, predictive coding has proved especially influential in guiding the development of neurobiologically-informed theories of online speech processing. These theoretical perspectives, which augment standard predictive coding formulations with insights gleaned from the predictive timing and active sensing literatures, are introduced next.

7.0.2 Predictive coding and predictive timing in speech perception

As alluded to in Chapter 3, predictive coding is an old idea that has become exceedingly influential over the past 20 years (Friston, 2018, 2019b). Given its generalised formulation under the free energy principle, predictive coding affords a powerful computational framework for modelling hierarchically-organised perceptual state dynamics within the brain (Bastos et al., 2012; Friston and Kiebel, 2009). One of the particularly appealing features of this framework with respect to the aforementioned debates in the language comprehension literature is its capacity to reconcile various competing conceptions of prediction under a single, unified scheme. On this view, the uncontroversially predictive aspects of higher-order language processing (e.g., anticipating the identity of an upcoming word), the more prosaic elements of speech perception (e.g., phoneme categorisation), and everything in between are explained as emergent properties of a hierarchical system that has evolved to optimise uncertainty over multiple temporal scales.

Predictive coding models are generally concerned with the problem of inferring the causes of an agent’s sensory states, and thereby resolving the contents of perceptual experience.

³Perhaps more surprising was the longstanding reluctance amongst members of certain linguistic traditions to accord prediction any significant role in language comprehension (DeLong et al., 2014; Huettig, 2015), despite early evidence to the contrary (e.g., Miller and Isard 1963; Tulving and Gold 1963).

The sentence processing literature is no exception to this tendency, replete as it is with paradigms that have been carefully crafted to induce expectations about the content of linguistic items. However, the expectations engendered by spoken language (and indeed, auditory stimuli more generally) aren't limited to content, but also pertain to temporal relations. This latter remark speaks to the notion of *predictive timing* ([Arnal and Giraud, 2012](#)), a generalisation of predictive coding that supplements inferences about the causes of sensory events with inferences about their temporal structure.

By extending the predictive coding scheme to incorporate temporal expectations, predictive timing brings predictive coding into contact with a large corpus of work on the temporal organisation of neural dynamics and behaviour (see, e.g., [Large and Jones 1999](#); [Nobre et al. 2007](#); [Nobre and van Ede 2018](#)). Tapping into this literature opens up exciting opportunities for mapping the computational machinery of predictive coding (and more generally, active inference) onto neurophysiological mechanisms in the brain. Of particular interest here is the influence of temporal prediction on speech perception as mediated via the interface of endogenous neural oscillations – a ubiquitous property of cortical networks that has been implicated in a variety of computational functions (see, e.g., [Buzsáki and Draguhn 2004](#); [Buzsáki 2010](#); [Engel et al. 2001](#); [Fries 2005, 2015](#); [Wang 2010](#)).

Neural oscillations are organised in a nested hierarchy reminiscent of the multi-scale temporal structure of natural language, making them a prime candidate in the search for the neurophysiological substrates of sentence processing. While distinct functional roles have been ascribed to various frequency bands ([Ghitza and Greenberg, 2009](#); [Ghitza, 2011](#); [Giraud and Poeppel, 2012](#); [Kösem and van Wassenhove, 2017](#); [Meyer, 2018](#)), most relevant from a predictive timing perspective is the encoding of (predominantly) syllabic information by low-frequency oscillations ($\sim 2\text{-}8$ Hz) that track the temporal modulation of the speech envelope ([Ahissar et al., 2001](#); [Abrams et al., 2008](#); [Doelling et al., 2014](#); [Ding and Simon, 2012](#); [Luo and Poeppel, 2007](#); [Pasley et al., 2012](#); [Pelle and Davis, 2012](#)). Evidence that speech becomes less intelligible as the quality of cortical speech tracking deteriorates ([Ahissar et al., 2001](#); [Gross et al., 2013](#); [Kerlin et al., 2010](#); [Luo and Poeppel, 2007](#); [Nourski et al., 2009](#); [Pelle et al., 2013](#)) supports the view that low-frequency oscillations are crucial for parsing continuous speech into discrete ‘packets’ of sub-lexical information that constitute the building blocks of higher-order linguistic structure ([Ghitza and Greenberg, 2009](#); [Ghitza, 2011, 2012, 2013](#); [Giraud and Poeppel, 2012](#)). Predictive timing finesses this view by positing that temporal expectations help to ‘tune’ low-frequency oscillations in to salient features of the acoustic stream, thereby improving the efficiency of speech encoding ([Arnal and Giraud, 2012](#)).

The involvement of predictive timing mechanisms in speech processing is supported by evidence that low-frequency oscillations are adaptively regulated to optimise the processing of rhythmic input (Schroeder and Lakatos, 2009; Morillon et al., 2015). The regular temporal patterning of (quasi-)periodic stimuli enables neural ensembles to synchronise their periods of high excitability with unfolding sequences of sensory events, a process often referred to as ‘entrainment’ (for reviews, see Lakatos et al. 2019; Haegens and Zion Golumbic 2018; Obleser and Kayser 2019). While such phenomena may in principle be induced in an entirely bottom-up fashion, nonhuman primate studies have provided compelling evidence of their regulation via top-down attention (Lakatos et al., 2008, 2009, 2016). Within the auditory cortex, this mechanism has been conceptualised as a ‘spectro-temporal filter’ (Lakatos et al. 2013) that selectively prioritises one acoustic stream of events at the expense of others (cf. *active sensing*; Schroeder and Lakatos 2009; Schroeder et al. 2010). Electrophysiological recordings in humans have provided tantalising evidence that analogous predictive mechanisms are recruited in the service of speech encoding (e.g., Mesgarani and Chang 2012; Zion Golumbic et al. 2013; see also Zoefel and VanRullen 2015).

7.0.3 The problem with predictive coding

The discussion thus far has concentrated on the influence of predictive coding on contemporary developments in the speech recognition and language comprehension literatures. While this represents a narrowing of focus relative to the overarching theme of active inference, it is not a surprising one; speech comprehension is after all firmly ensconced within the perceptual domain – quibbles about the involvement of motor activity in speech processing (e.g., Pickering and Clark 2014; Scott et al. 2009) and predictive timing (e.g., Morillon et al. 2014, 2015) aside, a significant part of any broader active inference account of speech comprehension is going to be founded on a perceptual inference scheme that recapitulates the hierarchical organisation of the cortical auditory system (cf. Bornkessel-Schlesewsky and Schlewsky 2013; Heilbron and Chait 2018; Rauschecker 1998; Rauschecker and Scott 2009). From this perspective, it is entirely understandable that predictive coding models have come to dominate the landscape.

It turns out, however, that predictive coding might not be as suited to the task of sentence comprehension as its influential standing within the literature would imply. The problem with predictive coding, as recently highlighted by Friston and colleagues (2021), is that it lacks a mechanism for arbitrating the most likely parsing of the speech stream amongst the many possible parsings it may support. This problem stems from the fact that the acoustic properties of continuous speech “show little respect for linguistic boundaries” (Pelle and Davis 2012, p. 2); contrary to written text, where consecutive

words are clearly demarcated by intervening spaces, the acoustic landmarks of speech do not reliably distinguish where one word ends and another begins. The ill-posed nature of word recognition in the context of online speech processing thus renders predictive coding schemes prone to erroneous inferences about sentence-level content.

Friston and colleagues (2021) propose to overcome this problem by augmenting predictive coding (or rather, an ‘amortised’ version of it – see Friston et al. 2021, for technical details) with a covert form of active inference. In this ‘active listening’ scheme, covert action pertains to the implicit placement of candidate word boundaries in accordance with acoustic cues availed by the speech envelope. This active segmentation process engenders a set of candidate partitions that are evaluated in conjunction with a generative model of word production. The winning partition is that sequence which maximises Bayesian model evidence (based on alternate inversions of the generative model) over its constituent words.

Characteristic of the active inference formalism, active listening entails a dialectic process whereby perception (word recognition) and action (boundary placement) are jointly functional to uncertainty reduction. Under this scheme, prior beliefs about the way words are generated enforce constraints on the segmentation of the acoustic stream, while the arbitration of competing segmentation options resolves uncertainty over lexical content. Moreover, simulations of active listening indicate that prior information plays a decisive role in determining the correct parsing of sentences heard under adverse conditions (Friston et al., 2021), an observation germane to the discussion of the empirical findings presented below.

7.0.4 The present manuscript

The manuscript that follows presents a study of sentence-level speech processing that investigates the effects of prior information on degraded speech comprehension. As discussed in the Introduction of the following manuscript, studies of degraded speech stimuli have rendered compelling evidence of the constructive nature of auditory perception, while also affording more specific support for a predictive coding account of speech recognition. However, most studies to date have focused on the processing of degraded words; those few studies to have investigated degraded sentence processing have rendered rather equivocal results with respect to predictive mechanisms.

This experiment attempted to overcome some of the methodological limitations of previous sentence-level studies, while also aiming to shed light on the involvement of covert action in sentence segmentation. To this end, we recorded high-density EEG while participants listened to degraded sentence stimuli, from which speech tracking and spectral

power estimates were derived. The results of this analysis are interpreted as furnishing preliminary evidence of the active engagement of attentional processes consistent with the active listening perspective on sentence segmentation. From a broader perspective, this finding provides further insight into the covert psychophysiological mechanisms underwriting the cognitive system's adaptation to environmental uncertainty.

Finding meaning in the noise: Expectations guide attention towards the content of degraded speech

Andrew W. Corcoran, Ricardo Perera, Matthieu Koroma,
Sid Kouider, Jakob Hohwy, Thomas Andrillon

Abstract

Neuroimaging studies of sub-lexical filling-in and word-level pop-out provide compelling evidence for the role of predictive mechanisms in speech comprehension. However, relatively little is known about the way such mechanisms contribute to the extraction of higher-level linguistic structure. This study leveraged the pop-out phenomenon (i.e. the dramatic improvement of degraded speech intelligibility following information about speech content) to investigate the neurophysiological correlates of sentence-level speech processing. We recorded 64-channel electroencephalograms from 19 adults while they rated the clarity (i.e. intelligibility) of sentences that had been transformed into noise-vocoded or sine-wave speech. Identical 10.5 s sentence stimuli were heard before and after a 4 s interval in which the participant was visually presented either the written version of the degraded sentence, written text from another (unheard) sentence, or no written information. Clarity ratings were significantly improved following the provision of correct sentence information only. This improvement in intelligibility was accompanied by a significant increase in cortical speech tracking, as operationalised by the quality of stimulus envelope reconstruction. Spectral power analysis further revealed that this effect was associated with the selective suppression of theta-band activity. Furthermore, delta- and alpha-band power were both enhanced following the provision of sentence information, irrespective of its content. These data extend previous studies of auditory recognition and continuous speech processing, providing novel insights into the predictive mechanisms underwriting sentence comprehension.

Keywords: EEG; Pop-out; Speech comprehension; Stimulus reconstruction; Predictive coding; Prior knowledge

7.1 Introduction

The ability to understand spoken language with rapidity and ease is a remarkable achievement of human cognition. In order to accomplish this feat, the brain must parse a continuous stream of acoustic modulations into a series of discrete units, and combine these units into meaningful segments of language. The transition from speech processing to language comprehension entails the online construction of syntactic and semantic relations subtending multiple temporal scales, and the flexible updating of these relations in accordance with novel information. In real-world settings, this already formidable task is often further complicated by the presence of competing speakers and ambient noise (Cherry, 1953), and the deviation of verbal expressions from the (sub)lexical representations onto which they must be mapped (Guediche et al., 2014). Such variability might derive from idiosyncrasies in speech production (e.g., accent, pathology), or corruption of the transmitted signal (e.g., filtering, temporal distortion; Mattys et al. 2012). This latter scenario – the problem of degraded speech comprehension – is the topic of the current study.

7.1.1 Perceptual restoration of degraded speech

Although the speech stream is characterised by exquisitely complex, fine-grained spectro-temporal structure, speech perception is remarkably robust to transient acoustic distortion. In a classic study by Warren (1970), participants listening to spoken sentences reported the illusory perception of phonemes that had been replaced by naturalistic sounds (coughs) or pure tones. This ‘restoration’ effect, which was not observed when phonemes were deleted without replacement, demonstrates the brain’s ability to construct stable speech percepts in the context of environmental noise.

Phoneme restoration (and auditory restoration/continuity effects more generally; see Bregman 1990; Petkov and Sutter 2011) can be construed as an auditory analogue of the ‘filling-in’ effects classically reported in the visual domain (Komatsu, 2006; Pessoa and De Weerd, 2003; Ramachandran and Gregory, 1991; Walls, 1954). In both sets of phenomena, an interrupted or occluded sensory pattern is unconsciously ‘interpolated’ on the basis of adjacent information to form a complete perceptual object or scene. Here, we use the term ‘perceptual filling-in’ to refer to a broad class of conscious states that includes instances of speech recognition in the absence (or severe degradation) of corresponding spectro-temporal properties in the acoustic environment (see also Shahin et al. 2009, 2012).

Beyond the restoration of individual phonemes, perceptual filling-in can be elicited in words and sentences that have been severely distorted. Using techniques such as noise-vocoding (Shannon et al., 1995) and sine-wave synthesis (Remez et al., 1981) to obliterate the fine-grained spectro-temporal features of natural speech (see Section 7.2.2 and Figure 7.1A), these highly degraded stimuli can be rendered intelligible by informing the listener of their original content (e.g., presenting an undistorted or written version of the utterance; Dehaene-Lambertz et al. 2005; Giraud et al. 2004). The provision of such information typically engenders an immediate improvement in the subjective clarity of the degraded utterance, a striking change in perceptual experience referred to as a ‘pop-out’ effect (Davis et al., 2005).

7.1.2 Neural mechanisms of perceptual filling-in

Filling-in phenomena have classically been adduced as evidence of the synthetic nature of speech perception (cf. Halle and Stevens 1959). Drawing on Rumelhart’s (1977) interactive schema theory, Samuel (1981) conceived of phonemic restoration effects as the product of interactions between top-down expectations based on prior knowledge and bottom-up acoustic-phonetic feature processing (cf. Marslen-Wilson 1975; Marslen-Wilson and Welsh 1978). This basic idea remains at the core of modern theorising about speech comprehension (e.g., Brodbeck and Simon 2020; Davis and Johnsrude 2007; Grossberg 2003; Grossberg and Kazerounian 2011; Heald and Nusbaum 2014; McClelland et al. 2006; Poeppel and Assaneo 2020), as exemplified by the widespread adoption of Bayesian predictive coding schemes (Bastos et al., 2012; Friston and Kiebel, 2009) in computational accounts of speech and language processing (e.g., Arnal and Giraud 2012; Bornkessel-Schlesewsky et al. 2015; Cross et al. 2018; Guediche et al. 2014; Lewis and Bastiaansen 2015; Meyer 2018; Poeppel et al. 2008; Poeppel and Monahan 2011; see also Hickok et al. 2011; Pickering and Garrod 2013).

From a predictive coding perspective, perceptual filling-in reflects unconscious inferences about the likely content of the auditory scene - the brain’s ‘best guess’ (i.e. prediction) about the sounds present in the environment (cf. Warren et al. 1972). A substantial corpus of neuroimaging evidence has accrued in support of this account. On the sub-lexical level, continuity illusions have been associated with a reduction in the auditory cortical activity typically evoked by vowel interruption, consistent with the top-down suppression of sensory prediction errors (Riecke et al., 2012; Shahin et al., 2012). Electroencephalographic (ECoG) recordings have further revealed that phonemic restoration is dependent on the top-down modulation of auditory cortical activity by the inferior frontal gyrus (Leonard et al., 2016), a region commonly implicated in functional magnetic resonance imaging (fMRI) studies of degraded speech comprehension (Binder et al.,

2004; Clos et al., 2014; Davis and Johnsrude, 2003; Eisner et al., 2010; Giraud et al., 2004; Hervais-Adelman et al., 2012; Lee and Noppeney, 2011; Obleser and Kotz, 2010; Shahin et al., 2009; Wild et al., 2012b; Zekveld et al., 2006).

On the word-level, several magnetoencephalography (MEG) and electroencephalography (EEG) studies have furnished convergent evidence that pop-out phenomena are similarly underwritten by predictive coding mechanisms. For example, auditory cortical responses to vocoded words are suppressed following congruent (but not incongruent) prior information (written text primes; Sohoglu et al. 2012; Sohoglu and Davis 2016). As would be expected under predictive coding, the magnitude of this effect is modulated by stimulus quality: While less-degraded speech evokes greater suppression when prior expectations are realised, neural activity is *enhanced* when expectations are violated (Sohoglu and Davis 2020; see also Blank and Davis 2016). This patterning is consistent with the view that the discrepancy between expectations and sensory inputs (i.e. *prediction errors*) depends on the quality (i.e. *precision*) of sensory input.

Evidence that predictive coding mechanisms support filling-in or pop-out phenomena at the sentence level is however relatively scarce. In line with the lexical priming studies mentioned above, fMRI data have indicated that sentence pop-out is associated with increased low-level sensory processing relative to clear speech (Tuennerhoff and Noppeney, 2016). ECoG data have further revealed that sentence pop-out is accompanied by the rapid tuning of auditory cortical ensembles to spectro-temporal speech features (Holdgraf et al., 2016). While the latter study was unable to confirm whether these effects were mediated by top-down processes, complementary EEG evidence suggests acoustic-phonemic encoding activity is suppressed (Di Liberto et al., 2018) – yet better aligned (or ‘entrained’) to the stimulus (Baltzell et al., 2017) – during the experience of sentence pop-out.

7.1.3 The current study

A notable difference between the word- and sentence-level studies surveyed above resides in the format of the stimuli used to instill prior expectations. The advantage of presenting written priors is that they convey abstract information about linguistic content via an unrelated sensory modality. As such, any influence of this input on acoustic-phonetic processing and speech recognition is likely to be mediated via top-down mechanisms (see Sohoglu et al. 2014; Wild et al. 2012a). By contrast, using the original version of the degraded utterance to disambiguate sentence content risks confounding changes in neural processing associated with perceptual pop-out with those induced by the auditory

system’s recent exposure to the clear sentence (cf. the induction of low-level ‘spectro-temporal priors’ over incoming sounds; [Holdgraf et al. 2016](#)). The current study sought to address this potential confound by manipulating prior knowledge of sentence content delivered in the form of written text. In this way, we were able to examine the neural correlates of sentence pop-out while holding the physical properties of the acoustic input constant.

To address this issue, we adopted a state-of-the-art modelling framework developed for the purpose of analysing electrophysiological measures of continuous speech encoding ([Crosse et al., 2016](#)). The general strategy here was to train a decoder on brain responses to continuous, undistorted speech, and use this decoder to reconstruct the acoustic envelope of degraded utterances from corresponding neural activity. The quality of stimulus reconstruction (as operationalised by its correlation with the acoustic envelope) was interpreted as a measure of cortical speech tracking, where better reconstruction was assumed to indicate a higher fidelity neural representation of the auditory stimulus. We complemented this approach with a time-frequency analysis over the low-frequency range (1-30 Hz) of the power spectrum. This additional analysis enabled us to assess whether differences in cortical speech tracking were associated with particular profiles of spectral activity.

7.2 Methods

7.2.1 Participants

Twenty-one native English-speaking adults were recruited to participate in this study. Of these, two were excluded due to faulty EEG recordings. The remaining sample comprised 8 females and 11 males aged 19 to 33 years ($M = 25.8$, $SD = 4.5$). All participants reported normal (or corrected-to-normal) vision and audition.

All participants provided written, informed consent, and were remunerated AU\$30 for their time. This protocol was approved by the Monash University Human Research Ethics Committee (Project ID: 10994).

7.2.2 Stimuli

A total of 80 pairs of English sentences were made. These pairs had similar grammatical structures and lengths (11.6 words on average). They were divided into 5 lists of 16 pairs (32 sentences per list). Each sentence was vocoded using Apple OS’s noise-to-speech command ‘say’ (voice = ‘Alex’, gender = male, sampling rate = 44.1kHz, rate =

200 words/minute). Each vocoded sentence was approximately 3.5 s long and was then concatenated three times to obtain audio files of ~ 10.5 s.

We then used publicly available scripts written for PRAAT ([Boersma and Weenink, 2011](#)) to turn clear speech into sine-wave speech (SWS) and noise-vocoded speech (NVS). In SWS, phonemes' formants are replaced by sinusoids at the same frequency, stripping the original clear speech from fine-grained temporal acoustic features and making SWS speech-like but unintelligible ([Remez et al., 1981](#)). In NVS, the amplitude of clear speech in a set of fixed logarithmically-spaced frequency bands (here, 6 bands) is used to modulate white-noise. This transformation preserves the temporal cues of the original signal but erases the spectral cues ([Shannon et al., 1995](#)). Consequently, SWS and NVS represent two complementary ways of degrading clear speech by removing fine-grained temporal cues (SWS) or spectral information (NVS; see [Figure 7.1A](#)).

The amplitude of the degraded speech was equalised across all sentences and the duration was adapted to a fixed 10.5 s duration using the VSOLA algorithm. In addition to these sentences, in the training session, we also played to participants an audiobook (Cat-Skin from Grimms' Fairy Tales, LibriVox) for a duration of 11'38". The properties of the speech (female voice, rate, etc.) were not modified except for the overall volume (same volume as SWS and NVS sentences). All auditory stimuli were delivered using the Psychtoolbox extension (v3.0.14; [Brainard 1997](#)) for Matlab R2018b (The MathWorks, Natick, MA, USA) running on Linux. The stimuli were played using speakers placed in front of the participant.

7.2.3 Procedure

Participants performed the experimental task while sitting at a desk with their head stabilised on a chinrest ~ 50 cm from the monitor. Following a 9-point eye tracker calibration, participants were instructed to actively attend to an audiobook (training) while maintaining fixation on a cross at the centre of the computer screen. They subsequently performed 6 blocks of 16 experimental trials each (test trials). Participants were instructed to maintain central fixation and refrain from excessive blinking while listening to the sentence presentations, but were permitted to blink and saccade outside these periods. Blocks were separated by self-paced breaks, with a recalibration of the eye-tracker prior to block 4. In total, the experimental procedure lasted approximately 75 min.

Each test trial started with the presentation of one noisy stimulus (NVS or SWS; 10.5 s long). Participants were then asked to rate the clarity of the noisy stimulus on a 4-point scale (1 = "I did not understand anything"; 2 = "I understood some of the sentence"; 3

= “I understood most of the sentence”; 4 = “I clearly understood everything”). Following this first clarity rating, participants were visually displayed either the corresponding written sentence (P+), or another sentence (P−), or no sentence (P0) for a fixed duration of 4 s. In all cases, the same noisy stimulus was presented a second time and participants were asked to rate the clarity of the stimulus using the same 4-point scale. Following this, when a sentence was visually displayed between the two presentations (P+ and P− conditions), participants were asked to indicate whether the displayed sentence corresponded to the noisy stimulus (Yes or No). A pause of 1.5 to 2 s (random jitter) was introduced before starting the next trial. See [Figure 7.1B](#) for a schematic illustration of the trial procedure.

Participants heard a total of 96 stimuli twice from all five lists. Each of these stimuli were presented in only one trial so that each stimulus is novel when presented the first time. This also means that the stimuli were heard exactly twice throughout the whole experiment and across all conditions. Each list was attributed to one condition (stimulus type: SWS or NVS; prior condition: P+, P− or P0). For the condition P+ and P−, the participants were exposed to only one sentence per pair from the corresponding Lists (see Section 7.2.2). This allowed us to present to participants, in the case of the P− condition, a sentence close to the stimulus played but different and never heard or seen before or after in the experiment. For the P0 condition, as no sentence is shown to the participant, we used both elements of each pair.

7.2.4 EEG acquisition and preprocessing

The EEG was continuously recorded during both the training (audiobook) and test trials (noisy speech) from 64 Ag/AgCl EasyCap mounted active electrodes using a BrainAmp system in conjunction with BrainVision Recorder (v1.21.0402; Brain Products GmbH, Gilching, Germany). Channels were digitised at a sampling rate of 500 Hz, with AFz serving as the ground electrode and FCz as the online reference.

Offline preprocessing was performed in MATLAB R2019b (v9.7.0.1319299) using custom-built scripts incorporating functions from the FieldTrip (v20200623; [Oostenveld et al. 2011](#)) and EEGLAB (v2019.1; [Delorme and Makeig 2004](#)) toolboxes. For the training data, EEG data were segmented in a single epoch starting 5 s before the start of the audiobook and ending 5 s after its end. For test trials, EEG data were segmented into 20 s epochs beginning 5 s before stimulus onset. All epochs were centred around 0 prior to high- and low-pass filtering (1 Hz and 125 Hz, respectively; two-pass 4th-order Butterworth filters). A notch (discrete Fourier transform) filter was also applied at 50 and 100 Hz to mitigate line noise.

For test trials, epoch and channel data were manually screened for excessive artefact using the ‘ft_rejectvisual’ function. A median 3 channels (range = 1–5) and 2 epochs (range = 0–5) were rejected per participant (note, an additional 5 trials were missing for one participant due to technical error). For training data, we performed only the channel rejection. Rejected channels were interpolated via the weighted neighbour approach as implemented in the ‘ft_channelrepair’ function (where channel neighbours were defined by triangulation).

Channels were re-referenced to the common average prior to independent component analysis (‘runica’ implementation of the logistic infomax ICA algorithm; [Bell and Sejnowski 1995](#)). Components were visually inspected and those identified as ocular ($Mdn. = 2$; range = 1–3), cardiac ($Mdn. = 0$, range = 0–1), or non-physiological ($Mdn. = 0$, range = 0–2) in origin were subtracted prior to backprojection. A separate ICA was run for the test and training data and artefact-related ICA components were identified separately.

7.2.5 Data analysis

7.2.5.1 Time-frequency decomposition

Preprocessed EEG data were re-referenced to the average of linked mastoids prior to time-frequency decomposition. Spectral power estimates were computed for epochs spanning -2 to 12 s relative to stimulus onset over a frequency range of 1 to 30 Hz (1 Hz increments) using the ‘ft_freqanalysis’ function (Hanning taper length = 1 s; 100 ms increments; see [Figure 7.2A](#)). To avoid potential confounding effects relating to stimulus repetition within each (1st and 2nd) presentation period, all subsequent analyses were limited to time-frequency estimates corresponding to the first sentence repetition (i.e. time-points spanning 0.5 to 3 s; first and last 0.5 s omitted to avoid spurious or confounding effects relating to stimulus onset/offset and spectral leakage). This decision also mitigates differences in task (dis)engagement over the course of the 2nd presentation period for trials in which pop-out was not experienced.

Channel-level time-frequency power estimates were averaged across time for each trial, and averaged across trials for each factorial combination of sentence type, prior condition, and presentation order. Averaged power estimates were then \log_{10} transformed and subjected to a nonparametric cluster-based permutation analysis ([Maris and Oostenveld, 2007](#)) as implemented in FieldTrip. Briefly, this procedure involves computing dependent samples t -tests across pairwise power estimates for each corresponding channel \times frequency bin, identifying t -values that exceed a specified alpha threshold (0.025,

two-tailed test), and clustering these samples into spatio-spectrally contiguous sets (minimum 2 neighbouring channels located within a 40 mm radius; average 3.9 neighbours per channel). T -values within each resolved cluster were then summed and the maximum value assessed against a Monte Carlo simulation-based reference (null) distribution (generated over 1000 random permutations).

To test the interaction of interest, the difference between 1st and 2nd presentation power estimates was contrasted across pairwise combinations of prior conditions for each sentence type. Clusters with a Monte Carlo p -value $< .05$ were deemed indicative of a significant difference between contrasts. Please note that this procedure only licences inferences about the existence of a statistically significant difference between contrasts; it does not permit the topographic or spectral localisation of such effects (see [Maris 2012](#); [Maris and Oostenveld 2007](#); [Sassenhagen and Draschkow 2019](#)). This caveat notwithstanding, the frequency bounds of the resolved clusters were used to inform the selection of frequency band limits in the subsequent linear mixed-effects analysis (Section 7.2.5.3).

7.2.5.2 Stimulus reconstruction

We used a stimulus reconstruction approach to estimate, from the EEG recordings, the quality of auditory processing. In particular, we focused on the reconstruction of the auditory envelope of the noisy speech from EEG recordings. Our rationale was that participants' ability to extract relevant cues from the noisy speech should be reflected in a better entrainment of EEG activity by the noisy speech and, therefore, as a better ability to reconstruct this envelope from EEG recordings. A similar approach was successfully applied to decode attention when participants are exposed to clear speech ([Legendre et al., 2019](#); [O'Sullivan et al., 2015](#)) or to reconstruct the envelope of NVS ([Di Liberto et al., 2018](#)).

We first extracted the acoustic envelope of the training and test stimuli in the 2-8 Hz band. This band was chosen for its correspondence with speech's syllabic rhythms and the robust entrainment of EEG oscillations with speech envelope observed in this frequency band ([Ding and Simon, 2014](#); [Giraud and Poeppel, 2012](#); [Pelle and Davis, 2012](#)). To do so, we ran the 10.5 s noisy speech (NVS and SWS) as well as the training stimulus through a peripheral auditory model using the standard Spectro-Temporal Excitation Pattern approach (STEP; [Leaver and Rauschecker 2010](#)). The stimuli were first resampled at 22.05 kHz and passed through a bandpass filter simulating outer and middle-ear preprocessing. Cochlear frequency analysis was then simulated by a bank of linear gammatone filters ($N = 128$ filters). Temporal integration was applied

on each filter output by applying half-wave rectification and a 100 Hz low-pass 2nd-order Butterworth filter. Next, square-root compression was applied to the smoothed signals and the power in each frequency band was log-transformed. Finally, the auditory envelope was computed by summing the envelope of the 128 gammatone filters and downsampled to 100 Hz.

For each presentation of the stimuli (training or test stimuli), we preprocessed the EEG recordings as follows. ICA-corrected epoched data were re-referenced to the average of all EEG electrodes and bandpass-filtered between 2 and 8 Hz using a two-pass Finite Impulse Response (FIR) filter and then resampled at 100 Hz. We trimmed the EEG epochs so that the start and end correspond to the start and end of stimulus presentation. We then used the Multivariate Temporal Response Function (mTRF) Toolbox for Matlab (v2.0; [Crosse et al. 2016](#)) to build a linear model between auditory and EEG signals from the training session (clear speech). By using an independent part of the experiment compared to test trials, and by using clear speech, we ensured that the model was not affected by our experimental design and represented normal speech processing. EEG data were shifted compared to the auditory envelope from 0 ms to 300 ms (31 time lags), which allows the integration of a broad range of EEG data to reconstruct each stimulus time point. The linear model was optimized to map the EEG signal from each electrode and time lag to the sound envelope. The obtained filter (matrix of weights: sensor \times time lags) was then used in the test trials to reconstruct the stimuli.

In the test trials, we used the model trained on clear speech to reconstruct the envelope of the noisy stimuli. This was done independently for each of the two presentations of the stimuli in each trial. Finally, the reconstructed envelope was compared to the envelope of the noisy stimulus played for this trial (NVS or SWS) by computing the Pearson’s correlation coefficient between the real and reconstructed envelope. We computed this coefficient for the three repetitions of the same sentence in each stimulus presentation. This coefficient (bounded between -1 and 1) was used as an index of the quality of the stimulus reconstruction. As per the time-frequency analysis, we focused on the first presentation of the sentence within a given trial and the first following the presentation of the correct (P+), incorrect (P−) or no (P0) visual sentence information.

7.2.5.3 Statistical analysis

Statistical analysis of trial-level subjective clarity ratings, frequency band power, and stimulus reconstruction scores was performed in R (v3.6.2; [R Core Team 2019](#)). Our general strategy for each analysis was to fit the appropriate mixed-effects model to the dependent variable of interest from the 2nd presentation, and regress these estimates

onto the corresponding estimate from the 1st presentation (including the 1st presentation estimate as a covariate essentially functions as a form of baseline correction; see [Alday 2019](#)). Additional independent variables were stimulus type (SWS, NVS), prior condition (P+, P-, P0), and the interaction between these factors, which were introduced into the model in that order. Model comparisons (log-likelihood) were performed using the ‘anova’ function to assess whether the additional complexity introduced by each new fixed effect term was merited by a sufficient improvement in model fit. Categorical variables (stimulus type, prior condition) were sum-to-zero coded (reference level = -1).

All mixed-effects models were fit with by-participant random intercepts. We attempted to fit maximal random effects structures for fixed effects of interest (i.e. stimulus type, prior condition) on this intercept ([Barr et al., 2013](#)); models that did not support this degree of complexity were reduced as reported in the results section. Random intercepts were also specified for sentence items in all models; EEG electrode channel locations were included as random intercepts in the spectral power models only (see [Liebherr et al. 2021](#), for a similar approach).

Subjective clarity ratings following the 2nd presentation were modelled as ordinal data using a (logit-linked) cumulative link mixed-effects model (i.e. proportional odds mixed model). This model was fit via the Laplace approximation using the ‘clmm’ function from the *ordinal* package ([Christensen, 2019](#)). No assumptions about the distance between cut-point thresholds were specified.

Linear mixed-effects models for frequency band power (averaged over time and frequency bins; first sentence iteration only) and stimulus reconstruction scores (first sentence iteration only) were fit using the ‘lmer’ function from the *lme4* package (v1.1-23; [Bates et al. 2015](#)). In addition to fixed effects described above (which were again introduced in a sequential fashion to enable model comparison), an ordered factor encoding the clarity rating on the 1st presentation was included as a covariate. Model diagnostics were assessed with the aid of the *performance* package (v0.5.0; [Lüdtke et al. 2021](#)).

The significance of main effect and interaction terms for each winning model was assessed using likelihood-ratio χ^2 tests from Type-II analysis-of-deviance tables obtained via the *RVAideMemoire* package (v0.9-79; [Hervé 2021](#)) for the cumulative link mixed-effects model; equivalent tables were obtained from the *car* package (v3.0-10; [Fox and Weisberg 2019](#)) for all linear mixed-effects models. Significant effects were disambiguated using post-hoc contrasts (Tukey corrected for multiple comparisons) obtained with *emmeans* (v1.5.1; [Lenth 2020](#)), which was also used to estimate marginal mean predictions for model visualisations. Model predictions and individual-level estimates were visualised with the aid of the *tidyverse* package (v1.3.0; [Wickham et al. 2019](#)).

7.3 Results

7.3.1 Correct prior information evokes perceptual pop-out

All participants performed at or near ceiling level in the discrimination of P+ from P− sentences ($M = 95.8\% \pm 1.3$ and $95.0\% \pm 1.4$ for SWS and NVS, respectively); these data are not subjected to formal analysis.

Individual- and group-level average clarity ratings are presented in [Figure 7.1C](#). Including prior condition within the cumulative link mixed-effects model yielded a significant improvement in fit ($\chi^2(9) = 1393$, $p < .001$); however, the model was not significantly improved by allowing prior condition to interact with stimulus type ($\chi^2(13) = 12.17$, $p = .514$). The main effect of stimulus type was significant ($\chi^2(1) = 11.71$, $p < .001$), indicating that clarity ratings tended to be higher following SWS than NVS sentences. A significant main effect was also observed for prior condition ($\chi^2(2) = 54.76$, $p < .001$). Post-hoc comparisons revealed a significant increase in clarity from P0 to P+ (z-ratio = 14.75, $p < .001$) and from P− to P+ (z-ratio = 15.99, $p < .001$), consistent with the experience of pop-out. Conversely, clarity levels did not significantly differ between P0 and P− (z-ratio = 1.05, $p = .547$).

7.3.2 Prior knowledge exerts frequency-specific effects on sentence processing

Grand-average time-frequency representations from the 2nd presentation period are displayed for each prior condition in [Figure 7.2A](#). Cluster-based permutation tests revealed significant differences in the average power across prior conditions for each stimulus type. Relative to the no-prior control (P0), receiving a correct prior (P+) resulted in a significant positive cluster (indicative of increased mean power) spanning 12 to 17 Hz in the SWS condition ($p = .006$), and 11 to 15 Hz in the NVS condition ($p = .005$). Similarly, receipt of an incorrect prior (P−) resulted in a significant positive cluster spanning 10 to 15 Hz in the SWS condition ($p = .002$); no significant clusters were identified for the corresponding NVS contrast. Topographies visualising the distribution of these clusters are presented in [Figure 7.2B](#).

On the basis of these results, a (high) alpha frequency band spanning 10 to 15 Hz was defined for mixed-effects modelling. Additional bands were also specified for the delta (1–3 Hz), theta (4–9 Hz), and beta (16–30 Hz) frequencies. In each set of nested model comparisons across the four frequency bands, the full model (i.e. including the

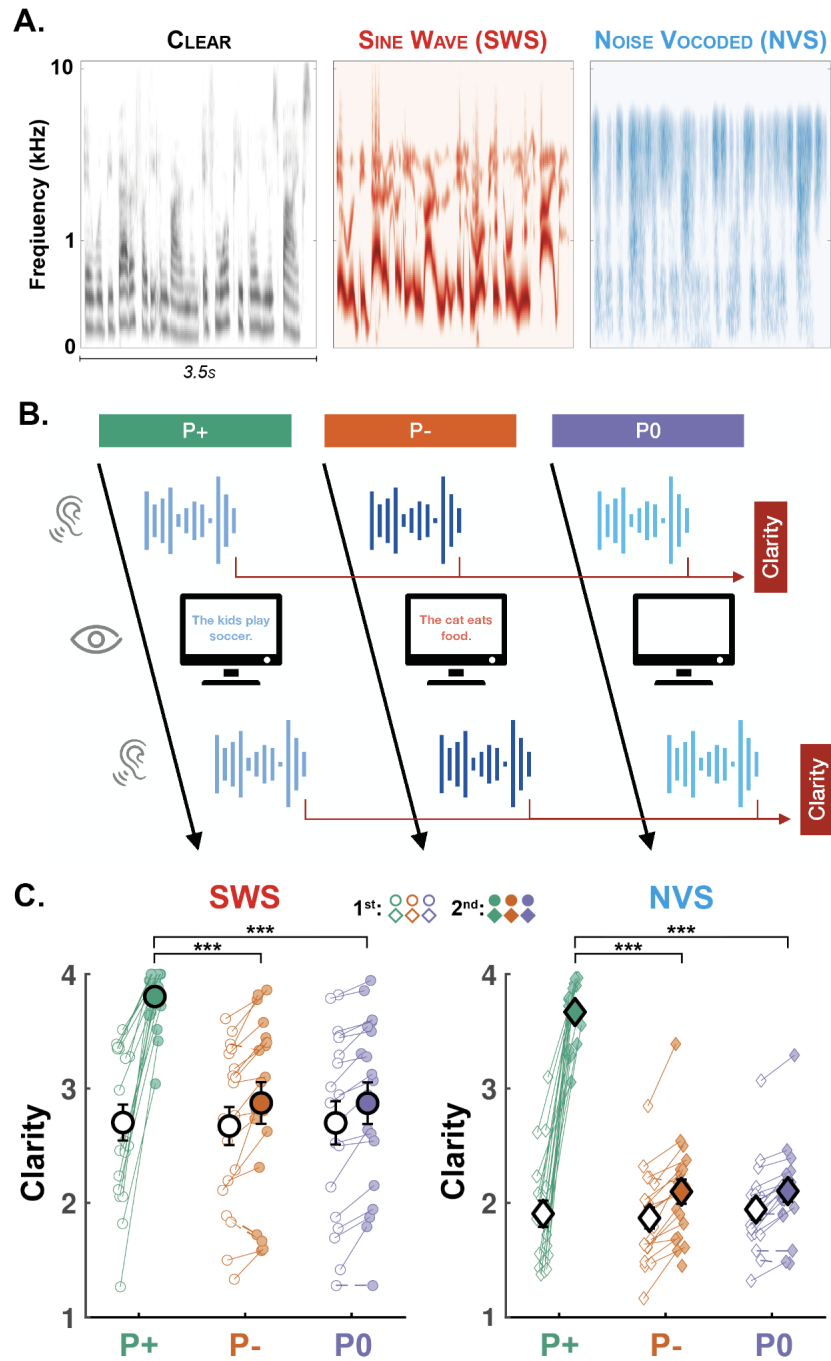


FIGURE 7.1: Experimental design and behavioural results. **A.** Cochlear representations of 3.5 s of clear speech (left), sine-wave speech (SWS; middle) and noise-vocoded speech (NVS; right). See Methods for details. **B.** In each trial, participants listened to two repetitions of the same noisy speech (SWS or NVS). The two presentations of the stimulus were interleaved with either (i) the corresponding written sentence (correct prior; P+), (ii) a different sentence (incorrect prior; P-), or (iii) no sentence (no prior; P0). Following each stimulus presentation, participants were asked to rate the subjective clarity of the stimulus. EEG was recorded throughout the task. **C.** Clarity ratings for SWS (left, circles) and NVS (right, diamonds) stimuli. Participants were asked to rate the stimuli after the 1st (unfilled circles and diamonds) and 2nd (filled circles and diamonds) presentations. Clarity ratings are averaged for each stimulus type and prior condition (P+: green, P-: orange, P0: purple). Individual data-points are shown with small circles and diamonds. Large circles and diamonds depict grand-averages ($N = 19$); error bars depict the standard error of the mean (SEM) across participants. Stars indicate significance levels of post-hoc contrasts across conditions (*** $p < .001$, ** $p < .01$, * $p < .05$).

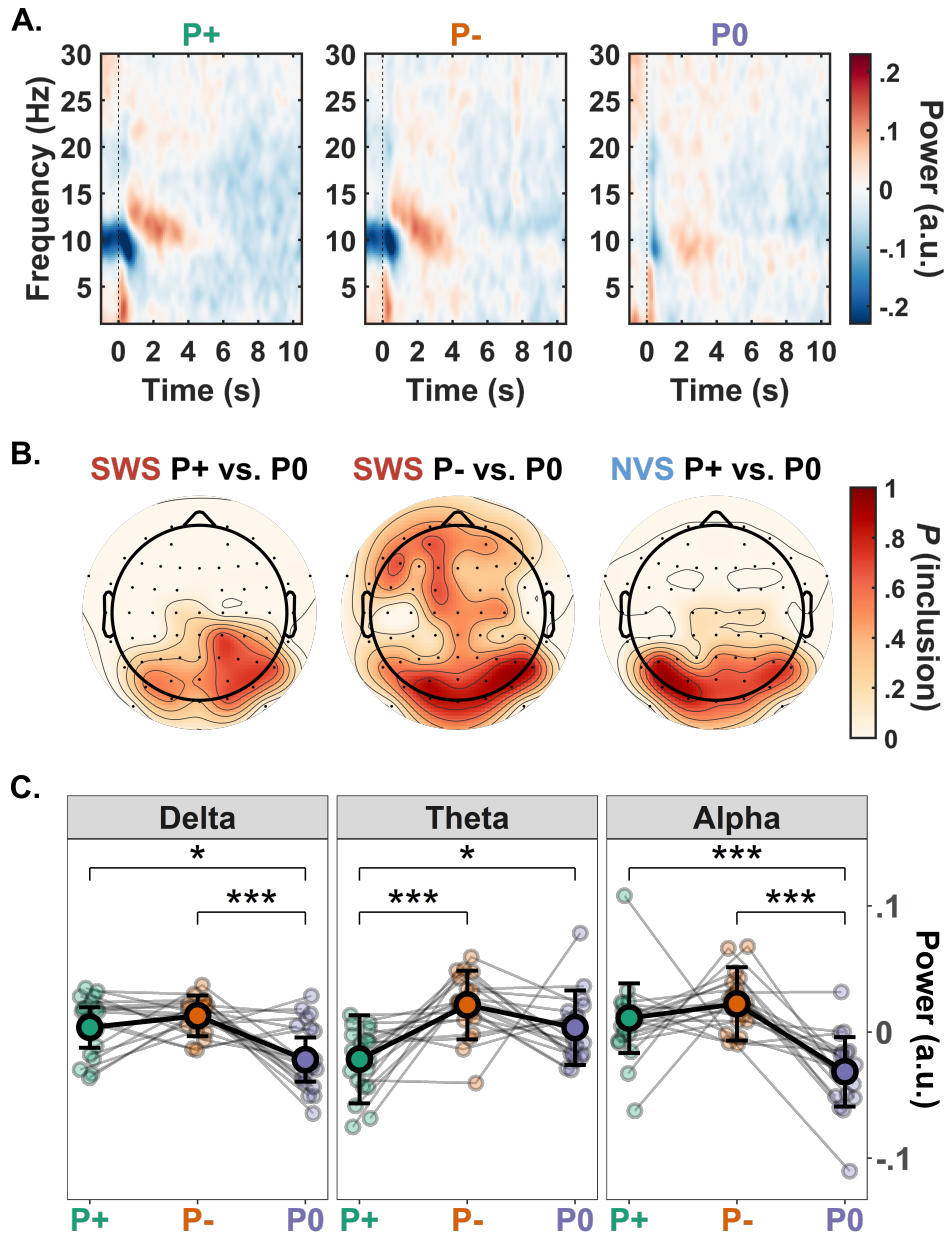


FIGURE 7.2: Time-frequency decomposition, cluster-based permutation analysis, and mixed-effects modelling of spectral power dynamics during sentence processing. **A.** Time-frequency representations depicting spectral power fluctuations over the course of the 2nd sentence presentation for each prior condition (P+, P-, P0). This period comprised 3 iterations of the same noisy stimulus (~3.5 s each). Spectral estimates were \log_{10} transformed, baseline-corrected with the time-average of estimates from the corresponding frequency bin in the 1st presentation period, and grand-averaged across stimulus types and subjects ($N = 19$). **B.** Topographic distribution of significant clusters identified via permutation analysis. Scale indicates the proportion of times an electrode at the corresponding location was included within the cluster (each plot averages over the 10-15 Hz band used in the alpha band analysis). **C.** Visualisation of linear mixed-effects models predicting mean spectral power during the first sentence iteration of the 2nd presentation period in the delta (1-3 Hz), theta (4-9 Hz), and alpha (10-15 Hz) frequency bands (beta-band model not visualised). Individual data-points are depicted with small circles. Large circles indicate estimated marginal means for the prior condition across the sample ($N = 19$); error bars show the 95% confidence interval across participants. Stars indicate the significance levels of post-hoc contrasts across condition levels ($***p < .001$, $**p < .01$, $*p < .05$). Note, model predictions have been mean-centred for the purposes of visualisation.

fixed effect and random slope interactions between stimulus type and prior condition) demonstrated significantly better fit than the reduced models.

Each frequency band model returned a significant main effect of prior condition, with the exception of the beta model (Delta: $\chi^2(2) = 16.51$, $p < .001$; Theta: $\chi^2(2) = 16.47$, $p < .001$; Alpha: $\chi^2(2) = 32.03$, $p < .001$; Beta: $\chi^2(2) = 0.19$, $p = .908$). Additionally, the main effect of stimulus type was significant in both the delta ($\chi^2(1) = 4.73$, $p = .030$) and theta ($\chi^2(1) = 8.17$, $p = .004$) models, on account of the increased spectral power evoked by NVS stimuli. The main effect of stimulus type was non-significant in both the alpha ($\chi^2(1) = 2.88$, $p = .090$) and beta models ($\chi^2(1) = 3.82$, $p = .051$), while the interaction between stimulus type and prior condition was non-significant across all models.

The estimated effects of prior condition on mean spectral power in the delta-, theta-, and alpha-bands are visualised on both the group- and the individual-level in [Figure 7.2C](#). Post-hoc comparisons revealed that delta power was significantly increased following P+ compared to P0 (z-ratio = 2.36, $p = .047$), and P- compared to P0 (z-ratio = 3.89, $p < .001$); the difference between P+ and P- was non-significant (z-ratio = 1.46, $p = .312$). Alpha power showed a similar pattern, whereby power was increased following P+ compared to P0 (z-ratio = 3.65, $p < .001$), and P- compared to P0 (z-ratio = 5.65, $p < .001$). Again, there was no difference in alpha power between P+ and P- (z-ratio = 0.98, $p = .588$). The theta model revealed a significant difference between P+ and P0 (z-ratio = 2.58, $p = .027$), and P+ and P- (z-ratio = 4.26, $p < .001$); the difference between P- and P0 was non-significant (z-ratio = 1.99, $p = .115$).

7.3.3 Correct prior information enhances stimulus reconstruction

Individual- and group-level average reconstruction coefficients are presented in [Figure 7.3A](#). The mixed-effects model for reconstruction scores revealed a significant main effect of stimulus type ($\chi^2(1) = 23.49$, $p < .001$), indicating that reconstruction scores, just as clarity ratings, are higher following SWS than NVS sentences. A significant main effect was also observed for prior condition ($\chi^2(2) = 15.98$, $p < .001$). Model comparisons indicated that models including interaction components (two-way interaction between prior condition and stimulus type, and three-way interaction between prior condition, stimulus type, and baseline reconstruction score) did not fit the data significantly better (all $p > .10$).

The main effect of condition was interrogated using post-hoc pairwise comparisons. These comparisons revealed that reconstruction scores were higher in the P+ compared to P0 (t-ratio = 3.66, $p < .001$) and P- (t-ratio = 3.22, $p = .004$) conditions, respectively.

Reconstruction scores did not significantly differ between P- and P0 (t-ratio = 0.44, $p = 0.90$).

Finally, we examined whether the stimulus reconstruction scores could predict the clarity of stimuli on the 2nd presentation above and beyond the stimulus condition (see [Figure 7.3B](#)). To do so we once again fitted cumulative link mixed-effects models on clarity ratings on the second presentation. We compared three different models: (1) a null model where stimulus type, prior condition and clarity ratings at the first presentation were used as fixed effects, (2) a second model adding reconstruction scores on the second presentation as fixed effects, and (3) a third model adding the three-way interaction between stimulus type, prior condition, and reconstruction scores. This third model revealed a significant three-way interaction between the predictors of interest ($\chi^2(2) = 8.58$, $p = .014$). However, a log-likelihood ratio test indicated that the inclusion of this interaction term did not significantly improve upon the fit achieved by the null model ($\chi^2(19) = 25.09$, $p = .158$).

7.4 Discussion

The present study set out to investigate the filling-in mechanisms that support perceptual pop-out while listening to noisy speech. To our knowledge, this is the first EEG study of degraded speech comprehension to use written information to induce sentence pop-out – thereby eliminating potential confounds stemming from previous auditory exposure to the clear version of the sentence. In our paradigm, we were able to manipulate the conscious experience of pop-out while holding the quantity and quality of acoustic stimulation constant across conditions. Our analysis of complementary EEG features (spectral power, speech envelope reconstruction) enabled us to identify distinct neurophysiological profiles in response to the manipulation of prior information. We interpret these results in relation to contemporary work on the frequency architecture of speech processing, and recent advances in the predictive coding/active inference literature.

7.4.1 Sentence pop-out is accompanied by enhanced stimulus reconstruction and theta suppression

In line with previous studies using written text primes to elicit the pop-out of degraded words, we observed marked increases in the reported clarity of degraded sentences when they were preceded by the correct (but not the incorrect or no) written sentence. Although noise-vocoded speech (NVS) was rated less clear on average than sine-wave

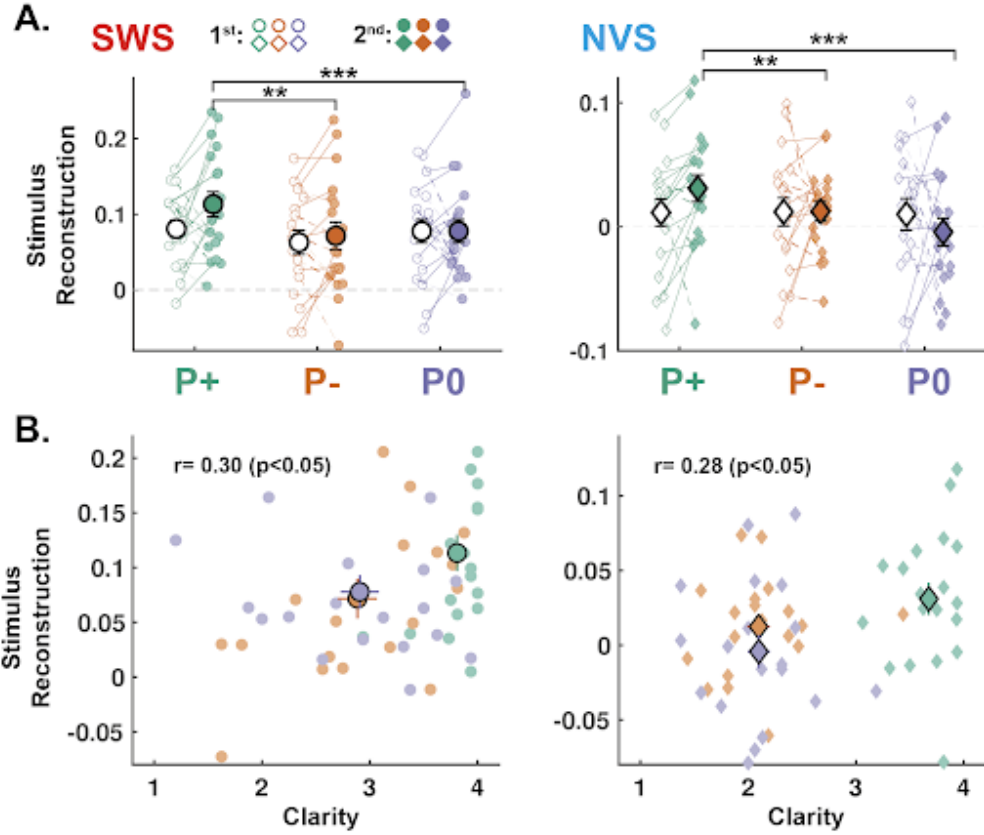


FIGURE 7.3: Correct priors improve stimulus reconstruction. **A.** The envelope of noisy speech was reconstructed from EEG recordings ($N = 19$, see Methods) and a stimulus reconstruction score was computed for the first 3.5 s (1st iteration of the sentence) of each stimulus presentation (1st: unfilled markers; 2nd: filled markers) and for the SWS (left, circles) and NVS (right, diamonds) stimuli separately. Reconstruction scores are averaged for each stimulus category and prior condition (P+: green, P-: orange, and P0: purple). Individual data-points are shown with small circles (SWS) and diamonds (NVS). The two average ratings of each participant and each category are connected with a continuous line if it increases from the 1st to 2nd presentation and a dashed line if it decreases. Large circles and diamonds show the average across the sample ($N = 19$); error bars show the standard error of the mean (SEM) across participants. Stars indicate the significance levels of post-hoc contrasts across condition levels ($***p < .001$, $**p < .01$, $*p < .05$). **B.** Correlation between clarity ratings and reconstruction scores on the 2nd presentation for SWS (left, circles) and NVS (right, diamonds). Individual data-points are shown with small circles (SWS) and diamonds (NVS). Large circles and diamonds show the average across the sample ($N = 19$); error bars show the standard error of the mean (SEM) across participants. The Pearson's correlation coefficient computed across conditions for the SWS and NVS is shown on each graph along with the associated p -value.

speech (SWS; although note the large degree of inter-individual variability apparent in [Figure 7.1C](#)), the pop-out effect was reliably obtained across both stimulus conditions.

The marked improvement in perceptual clarity following a correct written sentence was accompanied by two main electrophysiological correlates: (1) improved stimulus reconstruction, and (2) theta band (4–9 Hz) power suppression. The stimulus reconstruction finding is particularly compelling, because it suggests that amodal information obtained from the written sentence can modulate low-frequency EEG activity when one is listening to the sentence, such that the structure of the EEG signal more closely resembles that of the speech envelope. This finding is striking for at least two reasons. First, the participant never hears the undistorted version of the sentence at any point in the experiment – there is no low-level trace or prior ‘imprint’ that can be ‘reactivated’ by reading the sentence. Second, the decoding model used to reconstruct the speech envelope was trained on brain responses from an independent stimulus (i.e. audiobook narration), suggesting that improvements in stimulus reconstruction could be detected on the basis of generalisation from clear speech.

The enhancement of stimulus reconstruction quality in the P+ condition supports the notion that reconstruction quality is a reliable indicator of subjective speech clarity. This observation is consistent with previous reports that the quality of speech tracking (or entrainment) covaries with speech intelligibility ([Ahissar et al., 2001](#); [Luo and Poeppel, 2007](#); [Doelling et al., 2014](#); [Gross et al., 2013](#); [Peelle et al., 2013](#)). Such studies have tended to approach the relation between envelope tracking and intelligibility from a bottom-up perspective, reasoning that degraded speech becomes increasingly difficult to understand because of the progressive deterioration of speech tracking (cf. [Ding and Simon 2014](#); [Giraud and Poeppel 2012](#); [Peelle and Davis 2012](#)). Notably, however, the present findings would seem to resist such bottom-up explanations (see also [Kayser et al. 2015](#); [Lakatos et al. 2008](#); [Rimmele et al. 2015](#); [Zoefel and VanRullen 2015](#)), since our experimental manipulation modulated both stimulus reconstruction (tracking) and clarity (intelligibility) without any manipulation of acoustic features between the 1st and 2nd trial presentation (and holding repetition effects constant).

While a vast amount of work has been directed towards the role of theta phase dynamics in tracking, entraining, or parsing the speech stream, the significance of theta power modulations is less clear. As mentioned in the Introduction, a number of studies have reported suppression of auditory cortical activity during perceptual filling-in, some of which have directly implicated theta band suppression (e.g., [Riecke et al. 2009, 2012](#)). One might propose that being equipped with the correct prior information may help to ‘smooth over’ degraded spectro-temporal details in much the same way as phonemic restoration interpolates over transient noise mask. However, it’s not altogether clear

whether the neural mechanisms mediating filling-in on the timescale of phonemes or vowels would be able to support filling-in at the sentence level, and in the context of widespread (rather than localised) distortion.

A more domain-general perspective on the role of theta activity in speech processing inherits from the large corpus of evidence linking theta power to memory function (e.g., [Buzsáki 2002](#); [Buzsáki and Moser 2013](#); [Hanslmayr et al. 2016](#)). Of particular relevance from the perspective of sentence- and discourse-level language processing is the putative role of theta dynamics in the online integration and maintenance of representations over time ([Bastiaansen et al., 2010](#); [Cross et al., 2018](#); [Lam et al., 2016](#); [Piai et al., 2016](#)). In line with the predictive coding view discussed earlier, this line of research raises the possibility that knowledge of the correct sentence engenders lower theta power by virtue of furnishing the listener with an accurate hypothesis or model of the sensory input they are about to receive. Having a representation or template of the correct sentence already activated thus facilitates the online matching and integration process when input is received. In this way, the prior may reduce the computational burden on the retrieval and integration mechanisms indexed by theta activity.

7.4.2 Provision of prior information enhances delta- and alpha-band activity

Beyond the theta band, our spectral power analysis revealed elevated patterns of delta (1–3 Hz) and (high) alpha (10–15 Hz) band activity following the provision of prior information, relative to no information. Alpha band activity was also implicated in the cluster-based permutation analysis ([Figure 7.2A](#)), where topographic plots indicated a predominantly posterior scalp distribution. These findings would seem to have fewer precedents in the filling-in/pop-out literature reviewed above. However, past studies have identified posterior alpha band activity during degraded speech processing ([Obleser and Weisz, 2012](#)), albeit with some inconsistencies amongst word- and sentence-level studies ([McMahon et al. 2016](#); [Miles et al. 2017](#); see [Hauswald et al. 2020](#)). Such accounts often link alpha activity with effortful or attentive processing (e.g., [Dimitrijevic et al. 2019](#); [Obleser et al. 2012](#); [Wöstmann et al. 2015](#)), or relatedly, the inhibition of task-irrelevant networks ([Jensen and Mazaheri, 2010](#); [Klimesch et al., 2007](#); [Klimesch, 2012](#)), filtering-out of irrelevant information ([Kerlin et al., 2010](#); [Strauß et al., 2014](#); [Wöstmann et al., 2016, 2017](#)), or working memory demands ([Meyer, 2018](#); [Wilsch and Obleser, 2016](#)).

Arbitrating these various options is difficult, not least because they would seem to overlap with one another (e.g., directing the focus of attention is often effortful, and requires

one to ignore competing stimuli). However, a purely effortful account seems unlikely, given the nature of the pop-out effect; a similar argument might be mounted for working memory, unless incorrect priors are rejected and expelled from memory as rapidly as correct ones. Likewise, it seems unlikely that alpha (and delta) oscillations are elicited by purely sensory factors, which might be expected to affect all three conditions indiscriminately.

One speculative solution appeals to a recent innovation in the active inference literature. Friston and colleagues (2021) have proposed a model of sentence comprehension that augments the predictive coding architecture standardly appealed to in the filling-in literature with a covert mode of attentional processing. This mechanism serves to reduce uncertainty over the correct parsing of the speech stream by sampling different candidate segmentations and selecting the one that reduces the most prediction error. This view of covert attention might help to explain why we observe increased delta and alpha power following both sentence presentations, but not P0. That is, providing top-down priors about the possible location of word boundaries in the context of degraded speech might augment the automatic segmentation process, raising the possibility of new candidate boundaries to trial and evaluate. Moreover, this account is consistent with prior evidence implicating delta waves in the top-down selection and segmentation of words and phrases (Bonhage et al., 2017; Ding et al., 2016; Morillon et al., 2014).

7.5 Conclusion

This study provides the first compelling electrophysiological evidence that sentence pop-out is mediated by top-down mechanisms, adding to previous literature involving clean and/or word-level stimuli. By manipulating top-down prior information while holding bottom-up sensory information and prior stimulus exposures constant, we were able to show that correct sentence information and improved clarity were associated with better quality stimulus reconstruction. This finding is consistent with a predictive coding/active inference interpretation of envelope representation. Our spectral measures were indicative of distinct functional profiles, where theta activity was relatively reduced following correct sentences, while delta and alpha were increased following either sentence type. We interpret these findings as evidence that theta band activity indexes the efficiency of incremental sentence integration, while delta and alpha bands may be more characteristic of attentional sampling mechanisms. Future work should attempt to establish whether the delta-alpha pattern of activity can be teased apart, or whether these bands are functionally coupled during sentence segmentation.

8

Summary and concluding remarks

This closing chapter presents a brief rehearsal of the key moves that were made in each of the preceding six chapters, along with some reflections on the themes that emerged. I then conclude this thesis with some considerations for future research.

8.1 Looking back

Chapter 2 traced the historical development of the concept of allostasis through three of its major exponents: Sterling, McEwen, and Schulkin. While claims to the effect that allostasis supersedes classical notions of homeostasis appear to be overblown, the same can be said of arguments dismissing the former as a redundant reformulation of the latter. On the contrary, the predictive character of allostatic control appears to capture an important and distinctive dimension of biological regulation. Unfortunately, however, inconsistencies in the way allostasis has been conceptualised and deployed have propagated into the active (interoceptive) inference literature, resulting in confusion about the role of allostatic activity in the scheme of free energy minimisation.

In my review of the various assimilations of allostasis under the free energy principle, three overlapping conceptualisations of allostasis were identified, along with their most distinctive theoretical commitments. Of these, the ‘diachronic’ models of Pezzulo and

colleagues (2015) and Stephan and colleagues (2016) were identified as the most promising on the basis of their coherence with modern conceptions of allostasis in physiology and related fields, and their compliance with the formal mechanics of the free energy principle. The chapter closed with an appeal for greater clarity and precision in the way allostasis is conceptualised and integrated within the active inference framework – an objective taken up in the next chapter.

Chapter 3 sought to understand how biological regulation is conceptualised under the free energy principle, and whether the latter can be used to cast light on philosophical debates about the function and bounds of cognition. Noting certain affinities between active inference and Godfrey-Smith’s (1996) environmental complexity thesis, it was argued that the function of cognition is to assuage uncertainty by means of adaptive behaviour. However, active inference was argued to finesse this thesis by providing a framework that enables one to model the complexity of *internal* as well as external states, thereby furnishing the opportunity to characterise distinctive varieties of adaptive plasticity.

The crucial contribution of this chapter was the argument that certain forms of allostatic regulation may emerge from hierarchical generative models devoid of the capacity to evaluate expected free energy. While creatures embodying such architectures would be capable of inferring hidden dynamics spanning multiple timescales, they would not be capable of ‘counterfactual active inference’, a covert mode of action that entails the selective sampling of competing policies. The speculative proposal made on the basis of this argument was that cognition depends on counterfactual active inference; hence, creatures (or computational architectures) lacking this capacity are not apt to be described as cognitive systems.

Chapter 4 continued to pursue questions about the nature of cognition and its relation to physiological processes, this time from the perspective of embodied cognition. Active inference is often said to constitute an embodied account of cognitive processing and behaviour, and occasionally claimed to hold the key to resolving intractable disputes between proponents of cognitivism and those of embodiment. However, there appears to be very little consensus about which features of the framework justify such claims. Furthermore, much of the empirical work typically assumed to support embodied interpretations of active inference was argued to afford only weak evidence for embodiment (or indeed, only evidence for weak embodiment).

The positive side of this chapter proposed a novel theoretical hypothesis inspired by recent applications of active inference in the interoceptive and cognitive developmental domains. The key idea here was that rhythmic visceral feedback constitutes a vital stimulus for the early development and organisation of brain dynamics – essentially

‘sculpting’ or ‘carving out’ the fundamental structure of the generative model. This picture inverts the standard view of the brain as the central governor of bodily systems; instead, the brain is gradually ‘schooled’ to the point where it can begin to take on an ever-greater share of its regulatory responsibility. It also broadly recapitulates the hierarchical active inference scheme presented in Chapter 3, with homeostatic models embodied in brainstem reflexes gradually coming under the control of allostatic models embodied in subcortical centres, before cognitive capacities begin to emerge.

Chapter 5 marked the empirical turn towards psychophysiology, beginning with an experiment that integrated the early insights of the Laceys with contemporary active inference perspectives on covert action and heart-brain integration. A binocular-rivalry replay paradigm was used to manipulate the quality and quantity of uncertainty participants were exposed to while engaged in periods of attentive observation. These data were referenced against a baseline condition which involved minimal uncertainty in the visual domain, and made no demands on the focus of attention (i.e. eyes-closed resting-state).

As expected, the active deployment of attention to a visual scene that changed stochastically over time induced the classic ‘bradycardia of attention’; heart rate variability also declined. Crucially, these effects were amplified by the addition of perceptual ambiguity in the rivalry condition. Further analysis of time-resolved heart frequency estimates revealed additional sensitivity to stimulus novelty and perceptual alternation rate. These findings were interpreted as evidence that the cardiac correlates of active attention may be functional to uncertainty reduction in two complementary ways: (1) by reducing the impact of cardio-afferent feedback on exteroceptive processing (thus optimising signal-to-noise ratio in the visual domain); (2) by suppressing interoceptive prediction error caused by fluctuations in cardiac timing (thus optimising self-evidence).

Chapter 6 continued the investigation of cardiac dynamics as an expression of covert action, this time seeking to contextualise such activity in relation to spontaneous fluctuations in attentional state regulation. The Sustained Attention to Response Task variant of the Go/NoGo paradigm was employed to capture variability in overt behaviour relating to attentional fluctuations, which were subsequently mapped to changes in cardiac activity. Combining this paradigm with a thought-sampling technique meant that behavioural and cardiac dynamics could be further analysed in the context of attentional state reports.

While this study did not uncover evidence of any systematic relation between cardiac activity and self-reported attentional state, cardiac measures revealed interesting patterns of variation on very short and long timescales. The latter (time-on-task) effects were consistent with a few previous reports on this topic, and were interpreted as an

expression of covert ‘exploration’ in the context of prolonged sensory tedium. This interpretation thus affords an interesting counterpoint to the one presented in Chapter 5, suggesting cardiac dynamics can be enlisted to maximise or minimise surprise, depending on context. Attentional state also appeared to modulate the temporal co-ordination of cardiac and sensorimotor dynamics, although it would be premature to draw strong conclusions given the exploratory nature of this analysis.

Chapter 7 departed from the themes of cardiac regulation and brain-body integration, but continued to pursue the topic of covert attention in the familiar contexts of perceptual uncertainty and temporally-nested biological rhythms. The target phenomenon, perceptual filling-in, also reprised a theme that has flickered into view on a number of occasions, from the filtering of sensory inputs by hierarchical models (Chapter 3), to the elaboration of prototypical models during bistable perception (Chapter 5). This experiment investigated how the content of prior beliefs affects filling-in and covert attentional sampling mechanisms during the perception of degraded speech.

Stimulus reconstruction estimates revealed that the provision of accurate priors significantly improved the mapping between recorded brain activity and the speech envelope, consistent with the top-down modulation of sensory representations. This finding was complemented by evidence of theta power suppression relative to the incorrect and no prior conditions – an observation that fits nicely with recent work linking theta activity with online sentence integration, and previous data associating suppressed neural activity with sub-lexical filling-in and lexical pop-out. Finally, evidence that prior information induces higher delta and alpha power than no information was interpreted as a potential correlate of the active attentional processes hypothesised to underwrite the parsing of continuous speech.

Over the course of this series of theoretical and empirical chapters, I have tried to understand and explain a broad spectrum of biological, physiological, and cognitive phenomena from the perspective of active inference. In addition to this theory-driven approach, I have attempted to ground my inquiry in a sort of ‘historical consciousness’ (for want of a better phrase). What I mean by this is that I have endeavoured to situate my work, and that of many others, in the context of a rich and fecund history of ideas. In so doing, I have gradually been able to trace the roots of certain key concepts back to their origins, and observe how those ideas have undergone their own process of selection and adaptation over time. One might very well disagree with all manner of arguments put forth in this thesis, but if there is one abiding lesson to be drawn from these pages, it is surely the value that inheres in finding out how things came to be – and in finding one’s own place, in turn.

8.2 Looking ahead

But of course, one cannot spend all one’s time looking back. After all, these are exciting times for researchers interested in brain-body integration – especially for those like me who believe that nested cycles of neural, physiological, and behavioural activity offer vital clues about the relation between our biology and our minds. Interoception and allostasis are rapidly forcing their way to the forefront of cognitive scientific research (see, e.g., recent reviews by [Berntson and Khalsa 2021](#); [Chen et al. 2021](#); [Quigley et al. 2021](#)), driven by a precipitous increase in the number of studies investigating the impact of bodily states on cognition and neural function in health and disease ([Khalsa et al., 2018](#)). A similar trend has taken hold in the active inference community, as reflected by the increasing number of studies incorporating autonomic variables and allostatic mechanisms within their models.

Growing enthusiasm for interoceptive and allostatic modes of inference will undoubtedly lead to a proliferation of new perspectives, interpretations, and hypotheses. While abundance and plurality are to be welcomed as stimulants for inquiry, critical dialogue will be necessary to ensure the field does not succumb to the kind of fragmented (and at times, confused) theorising discussed in Chapter 2. It would be a mistake, for instance, to succumb to the temptation to characterise *all* forms of prospection, decision-making, action-planning, etc. as allostatic in character – a position that ultimately collapses any meaningful distinction between allostasis and cognition (cf. [Kiverstein and Sims 2021](#)). While I have sought to emphasise the deep continuity between physiological and cognitive forms of control, I have maintained that allostasis refers to a subset of adaptive capacities within the broader scheme of active inference. It’s not clear to me what theoretical insights or advantages are to be gained by generalising allostasis to the point that it becomes tantamount to policy selection – indeed, doing so seems to sacrifice a distinctive concept in exchange for a mere synonym.

Rather than broadening allostasis beyond its already ample scope, future work might profit from attempting to model and understand particular forms of biological regulation in their complexity. This could mean building active inference architectures that model interactions between multiple physiological cascades and feedback loops, or that evince a broad repertoire of context-specific response profiles in a single modality (cf. Chapter 5 and Chapter 6). Such work might also consider the particular substrates and biophysical conditions that are necessary (or sufficient) to implement prediction error minimisation, potentially highlighting certain forms of biological activity that challenge current formulations of active inference. Further, simulated agents designed to emulate autonomic and behavioural adaptations through active inference (e.g., [Tschantz et al.](#)

2021) might help to adjudicate theoretical claims about the formal distinctions between different varieties of adaptive behaviour (such as those discussed in Chapter 3).

Beyond the complexification of internal dynamics, active inference models might also be subjected to increasingly more complex external dynamics. This idea naturally follows from the analysis presented in Chapter 3, where environmental complexity (more specifically, different regimes of uncertainty) was posited as a stimulus for adaptation across multiple timescales. Of particular interest from the perspective of developmental biology, and building on the ideas presented in Chapter 4, future work might attempt to model the influence of external stimuli on self-organisation during embryonic, foetal, and infant development. (Note that prenatal development presents an especially interesting (and complex) situation from the perspective of allostasis and embodiment, whereby the boundaries between internal and external, self and other, and body and world are uniquely blurred; Ciaunica et al. 2021; Kingma 2019). Away from computational modelling, organoid technologies (Di Lullo and Kriegstein, 2017; Lancaster et al., 2013; Qian et al., 2019) may afford unprecedented opportunities for interrogating the role of embodiment in brain development. Such *in vitro* models could also provide a unique platform for testing and refining the visceral afferent training hypothesis.

With respect to the empirical findings presented in this thesis, there are a multitude of directions in which this work could be taken. I shall limit myself to three broad remarks, each of which are variations on my earlier appeal to embrace biological phenomena in their complexity and diversity. The first of these concerns modality. Much like the Laceys' psychophysiological heyday, recent cognitive neuroscientific interest in brain-body integration has mostly been an affair of the heart. Although this picture is starting to change, the impact of other physiological oscillations (e.g., gastric, respiratory cycles) on neural activity remains poorly understood. As interest in neuro-visceral communication diversifies beyond the brain-heart axis, I hope to see the emergence of multimodal experimental designs investigating the integrated co-ordination of allostatic activity across multiple interoceptive domains.

In addition to broadening empirical horizons beyond single dimensions of allostatic regulation, future work will also profit from attending to the inherent variability of such processes both within and between individuals. Individual differences in interoceptive sensitivity and autonomic variability have received considerable attention in recent years, featuring prominently in the search for biomarkers and endophenotypes (and potential targets for intervention; Bonaz et al. 2021). Indeed, the concept of allostasis was in some sense born out of a deep concern for variance across individuals (or groups thereof), and the consequences of such differences for health and disease (Sterling, 2020). As

briefly alluded to at the end of Chapter 5, there is much scope to explore how allostatic responses to uncertainty vary between individuals, particularly in the context of (psycho)pathology. This work might be productively combined with formal modelling techniques (Petzschner et al., 2021), including those being developed for the purposes of computational phenotyping (Friston et al., 2014; Patzelt et al., 2018).

Third and finally, a comment on the topic of affective experience. As mentioned in Chapter 1 and elaborated in Chapter 4, a great deal of interoceptive inference research deals with the genesis of subjective feeling states and their influence over exteroception and action. It might have come as a surprise, then, that I chose to orient my project away from this domain. One motivation for this decision was purely pragmatic: One can only do so much; better to explore the patch-less-visited than the one well-sampled. However, another motivation was more strategic: As articulated in Chapter 4, it seems to me that the strongest claims for the influence of visceral afferent feedback on cognitive activity are licensed by experiments that demonstrate its effects without manipulating bodily or affective experience.

That being said, manipulations that are not designed to induce emotional states or alter bodily percepts do not guarantee any sort of ‘affective neutrality’. Indeed, one might justifiably contend that the perceptual uncertainty experienced in binocular rivalry or degraded speech processing is aversive to some degree. Similarly, spontaneous fluctuations in attentional engagement, boredom, and arousal during the performance of a tedious task might plausibly covary with a more-or-less pleasant affective state. Thus, while I hold that this strategy provides a productive alternative to more traditional, direct modes of interoceptive perturbation, the ability to cleanly control or partial-out such influences is inherently challenging, and should not be over-stated. Instead, future work should seek to build a complementary picture that integrates the effects of interoceptive-, affective-, and bodily-state dynamics with more ‘covert’ methods of perturbing and measuring allostatic dynamics.

The interoceptive roots of affect and emotion brings one final, formidable challenge into view: the so-called ‘hard problem’ of consciousness (Chalmers, 1995). It is perhaps no accident that the increased prevalence of interoceptive inference, affective experience, and allostatic regulation within the active inference literature has coincided with increased interest in the nature of consciousness – a trend perhaps best exemplified by a subtle shift in Friston’s own language over the past few years, in which talk of self-organising systems has gradually given way to contemplation about *sentient* systems.

Whether or not the free energy principle possesses the necessary conceptual resources to solve the hard problem – or to reveal it as ill-posed (see, e.g., Hohwy 2021) – will no doubt be a topic of intense debate in the years to come. Considering its wide-ranging

explanatory ambitions, it would be surprising to discover that the free energy principle has nothing to say about the mystery of consciousness – as if subjective experience somehow inhabits an autonomous plane unto itself, free from the fundamental dynamics supposed to govern life in all its complexity. It would also seem odd to me if allostasis had no role to play in any such account; on the contrary, if the emergence of consciousness follows from some biological imperative to survive, it seems eminently plausible to me that the capacity to leverage one’s sensory experiences in the service of long-term viability occupies a central place in this story. As for the details of that story... those will have to wait for another day.





Supplementary materials

This appendix comprises the supplementary materials published alongside [Corcoran et al. \(2021\)](#) (Chapter 5).

Supplementary materials

Be still my heart: Cardiac regulation as a mode of uncertainty reduction

Andrew W. Corcoran 

Vaughan G. Macefield 

Jakob Hohwy 

Correspondence: andrew.corcoran1@monash.edu

List of Tables

S1	Stimulus identity	2
S2	Stimulus selection criteria	2
S3	LMM for switch rate	3
S4	GAMM for pulse amplitude	4
S5	GAMMs for skin potentials	5
S6	GAMMs for heart frequency (mean-centered)	7

List of Figures

S1	Instantaneous pulse amplitude by condition	4
S2	Skin potential amplitude by trial	6
S3	Instantaneous heart frequency by condition (mean-centered)	8
S4	Instantaneous heart frequency by trial (mean-centered)	9
S5	Instantaneous heart frequency by switch rate (mean-centered)	10

1 Materials: Face and house stimuli

The identity of each stimulus pair is documented in Table S1. Each stimulus was sampled from a subset of images satisfying the rating criteria detailed in Table S2.

Table S1: Stimulus identity.

	Face	House
Practice	CFD-WF-208-068-N	House84
Block 1	CFD-AF-216-106-N	House11
Block 2	CFD-AF-236-145-N	House20
Block 3	CFD-AM-215-120-N	House3
Block 4	CFD-AM-218-085-N	House33
Block 5	CFD-WF-011-022-N	House64
Block 6	CFD-WF-033-002-N	House82
Block 7	CFD-WM-204-031-N	House89
Block 8	CFD-WM-033-025-N	House86

Table S2: Stimulus selection criteria.

Face	House
Afraid < 3	Face-Likeness > 4
Angry < 3	Typicality > 4
Attractive 3 – 5	
Disgusted < 3	
Happy < 3	
Prototypic > 3	
Sad < 3	
Surprised < 3	
Threatening < 3	
Unusual < 2	

2 Results

The model summary for the linear mixed-effects model for switch rate is presented in Table S3. Model summaries for generalised additive mixed-effects models for instantaneous pulse and skin potential amplitude are presented in Tables S4 and S5. Summaries for mean-centered (but not z-normalised) instantaneous heart frequency data are presented in Table S6. Model visualisations are presented in Figures S1 (pulse amplitude), S2 (skin potentials), and S3-S5 (mean-centered heart frequency).

Table S3: Linear mixed-effects model for switch rate.

<i>Predictors</i>	<i>Estimate</i>	<i>S.E.</i>	<i>t</i>	<i>p</i>
(Intercept)	0.16	0.01	12.37	<.001
Condition	0.00	0.00	-1.65	.100
scale(IBM)	0.01	0.01	1.18	.240
scale(HRV)	0.00	0.00	-0.41	.681
Condition:scale(IBM)	0.00	0.00	-1.82	.069
Condition:scale(HRV)	0.00	0.00	-0.10	.922
<i>Smoothers</i>	<i>e.d.f.</i>	<i>d.f.</i>	<i>F</i>	<i>p</i>
s(ID _{subj})	36.70	38	33.32	<.001
s(ID _{stim})	10.57	15	2.24	<.001
Adjusted R^2	.63			
Observations	1244			

Table S4: Generalised additive mixed-effects model for instantaneous pulse amplitude.

<i>Predictors</i>	<i>Estimate</i>	<i>S.E.</i>	<i>t</i>	<i>p</i>
(Intercept)	-0.32	0.29	-1.10	.273
Replay	0.02	0.02	1.11	.265
<i>Smoothers</i>	<i>e.d.f.</i>	<i>d.f.</i>	<i>F</i>	<i>p</i>
s(Time):Rivalry	4.59	5.71	1.80	.093
s(Time):Replay	2.59	3.15	1.34	.270
s(Time, ID _{subj})	231.75	350.00	31.61	<. .001
s(Time, ID _{stim})	70.68	143.00	1.33	<. .001
Adjusted R^2	.45			
Observations	291993			

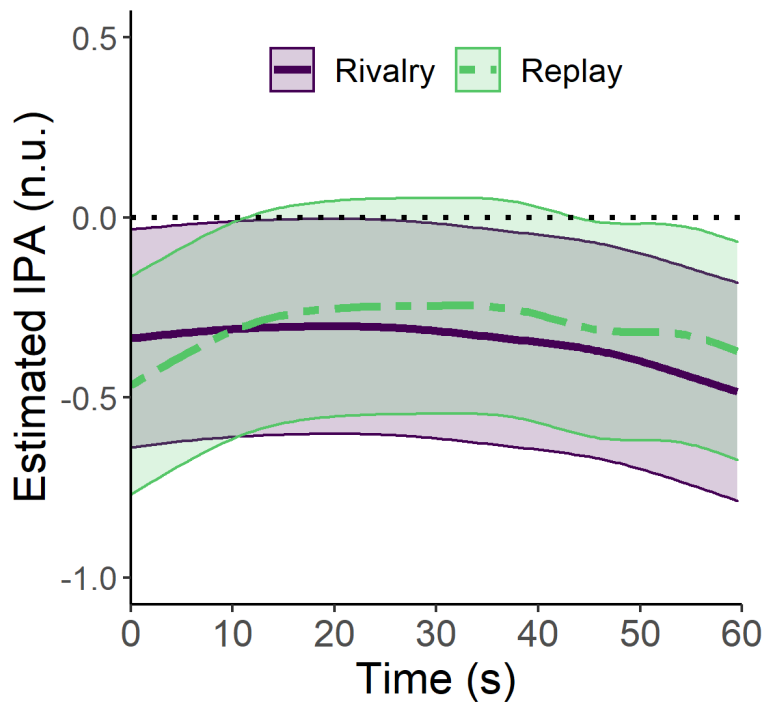


Figure S1: Time-resolved estimates of instantaneous pulse amplitude (IPA) across conditions. Dotted line indicates mean IPA during resting-state; negative IPA estimates represent decreased amplitude relative to this baseline. Shading indicates standard error of the mean. *n.u.* = normalised units.

Table S5: Generalised additive mixed-effects models for skin potential amplitude.

SPA by Condition					SPA by Trial				
Predictors	Estimate	S.E.	t	p	Predictors	Estimate	S.E.	t	p
(Intercept)	-0.06	0.08	-0.70	.484	(Intercept)	-0.06	0.07	-0.79	.432
Replay	0.04	0.01	3.50	<. .001	Rivalry1	-0.16	0.02	-9.30	<. .001
					Replay1	0.00	0.02	-0.14	.888
					Rivalry2	0.09	0.02	5.17	<. .001
Smoother					Smoother				
	e.d.f.	d.f.	F	p		e.d.f.	d.f.	F	p
s(Time):Rivalry	3.79	4.65	1.97	.091	s(Time):Rivalry1	2.62	3.27	2.21	.079
s(Time):Replay	1.59	1.77	0.32	.654	s(Time):Replay1	1.33	1.56	0.17	.706
					s(Time):Rivalry2	1.01	1.02	0.51	.481
					s(Time):Replay2	3.36	4.19	1.42	.247
s(Time, ID _{subj})	277.66	350.00	12.49	<. .001	s(Time, ID _{subj})	278.32	350.00	12.57	<. .001
s(Time, ID _{stim})	57.75	143.00	1.30	<. .001	s(Time, ID _{stim})	58.67	143.00	1.34	<. .001
Adjusted R^2	.11				Adjusted R^2	.11			
Observations	292317				Observations	292317			

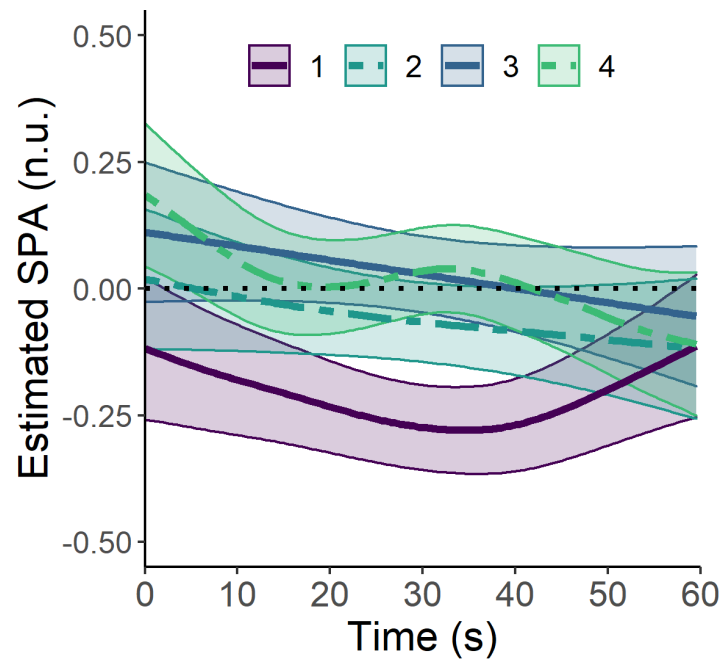


Figure S2: Time-resolved estimates of skin potential amplitude (SPA) across trials. Dotted line indicates mean SPA during resting-state; negative SPA estimates represent decreased voltage relative to this baseline. Shading indicates standard error of the mean. *n.u.* = normalised units.

Table S6: Generalised additive mixed-effects models for instantaneous heart frequency (mean-centered).

IHF by Condition					IHF by Trial				
Predictors	Estimate	S.E.	t	p	Predictors	Estimate	S.E.	t	p
(Intercept)	-0.04	0.01	-3.31	<.001	(Intercept)	-0.04	0.01	-3.36	<.001
Replay	0.01	0.00	9.85	<.001	Rivalry1	-0.02	0.00	-12.79	<.001
					Replay1	0.00	0.00	2.98	.003
					Rivalry2	0.00	0.00	1.40	.163
Smoothers					Smoothers				
	e.d.f.	d.f.	F	p		e.d.f.	d.f.	F	p
s(Time):Rivalry	8.21	8.51	29.62	<.001	s(Time):Rivalry1	7.70	8.34	12.77	<.001
s(Time):Replay	8.02	8.37	21.15	<.001	s(Time):Replay1	7.78	8.41	14.32	<.001
					s(Time):Rivalry2	8.48	8.80	29.71	<.001
					s(Time):Replay2	7.21	7.98	15.43	<.001
s(Time, ID _{subj})	284.92	350.00	29.61	<.001	s(Time, ID _{subj})	284.41	350.00	29.97	<.001
s(Time, ID _{stim})	108.64	143.00	3.76	<.001	s(Time, ID _{stim})	107.82	143.00	3.68	<.001
Adjusted R^2	.41				Adjusted R^2	.41			
Observations	293729				Observations	293729			

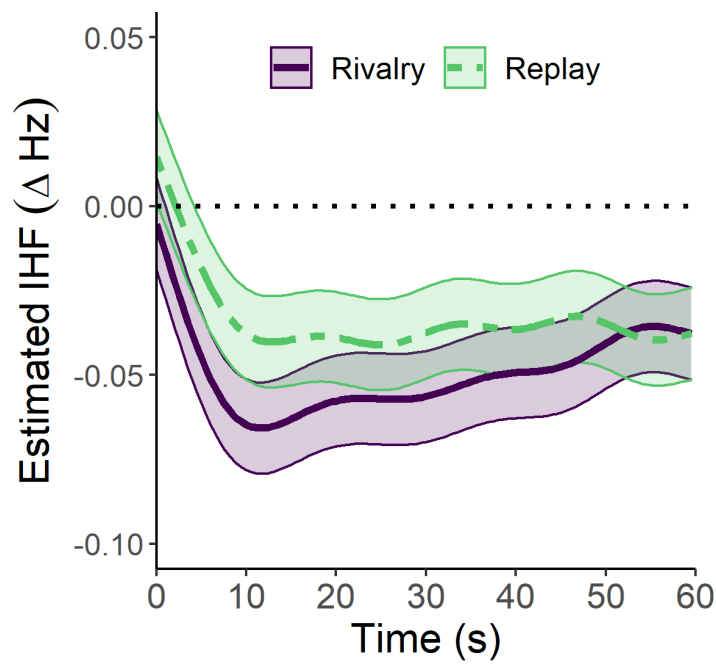


Figure S3: *Time-resolved estimates of instantaneous heart frequency (IHF) change across conditions.* Dotted line indicates mean IHF during resting-state; negative estimates represent a decrease in IHF relative to this baseline (i.e. cardiac deceleration). Shading indicates standard error of the mean.

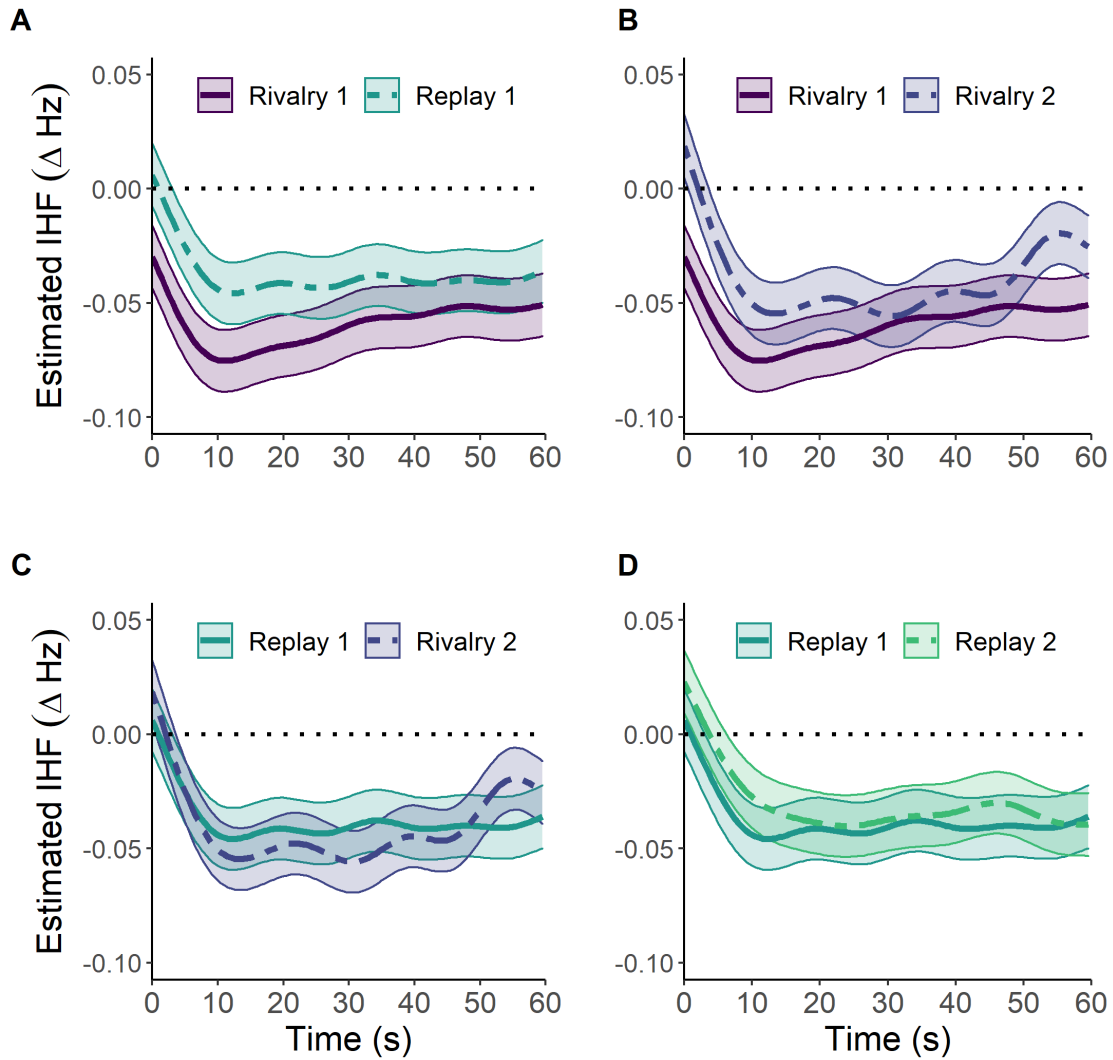


Figure S4: Pairwise comparisons of instantaneous heart frequency (IHF) change across trials. Dotted lines indicate mean IHF during resting-state; broken lines depict the later trial within each pair. Negative estimates represent a decrease in IHF relative to baseline (i.e. cardiac deceleration). Shading indicates standard error of the mean.

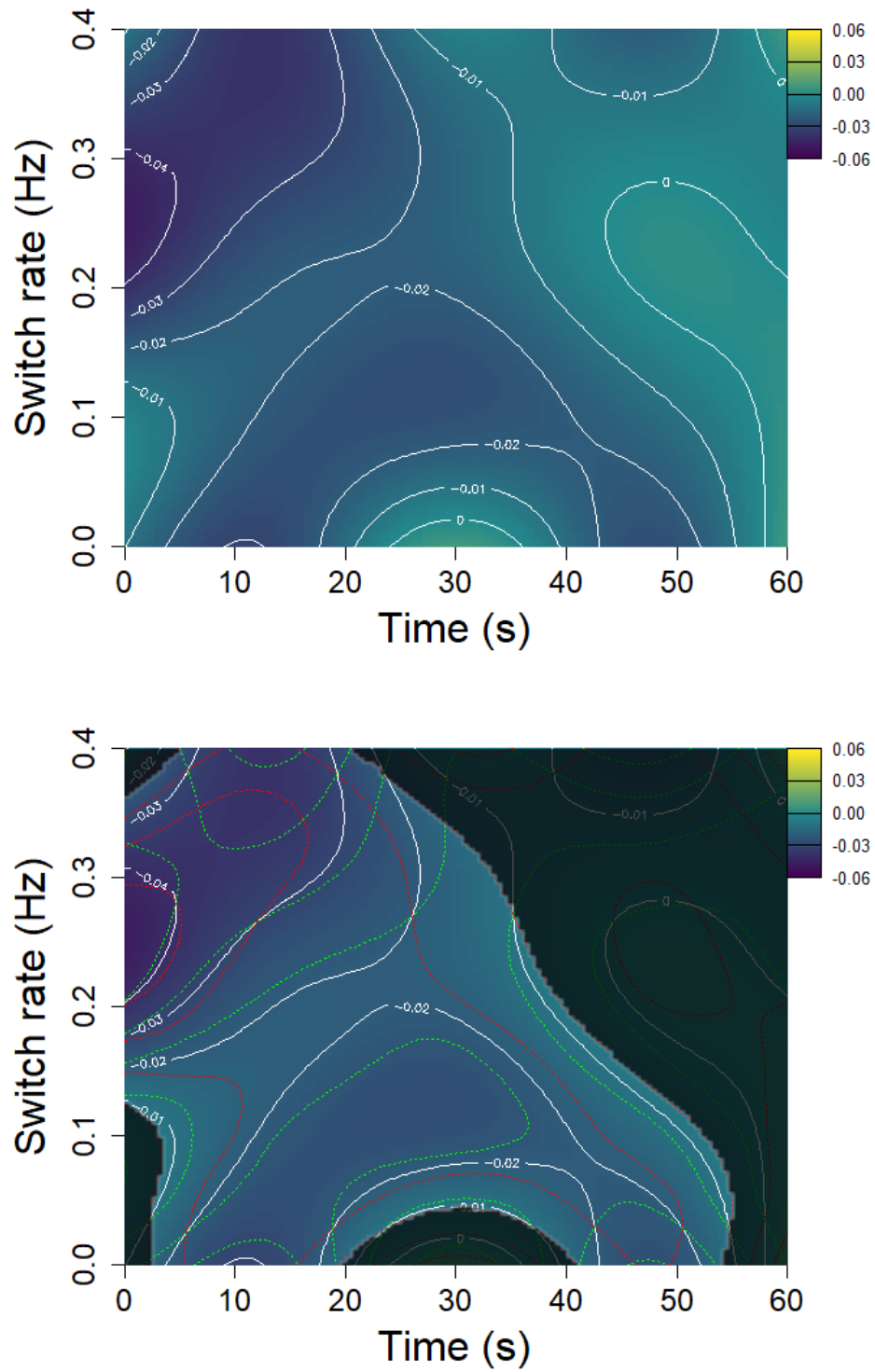


Figure S5: *Modulation of instantaneous heart frequency (IHF) by switch rate.* Top contour plot depicts differences in IHF (Hz) between conditions as a function of trial time and perceptual switch rate (green indicates no difference between conditions; darker blue indicates greater deceleration in rivalry). Bottom contour plot masks regions of non-significant difference and includes 95% confidence intervals for contours (red and green dotted lines).

Bibliography

- Abrams, D. A., Nicol, T., Zecker, S., and Kraus, N. (2008). Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *Journal of Neuroscience*, 28(15):3958–3965.
- Adams, F. and Aizawa, K. (2008). *The bounds of cognition*. Malden, MA: Blackwell.
- Adelhöfer, N., Schreiter, M. L., and Beste, C. (2020). Cardiac cycle gated cognitive-emotional control in superior frontal cortices. *NeuroImage*, 222:117275.
- Adler, D., Herbelin, B., Similowski, T., and Blanke, O. (2014). Breathing and sense of self: Visuo-respiratory conflicts alter body self-consciousness. *Respiratory Physiology & Neurobiology*, 203:68–74.
- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, 98(23):13367–13372.
- Ainley, V., Apps, M. A. J., Fotopoulou, A., and Tsakiris, M. (2016). ‘bodily precision’: A predictive coding account of individual differences in interoceptive accuracy. *Philosophical Transactions of the Royal Society B*, 371(20160003):1–9.
- Aizawa, K. (2007). Understanding the embodiment of perception. *Journal of Philosophy*, 104(1):5–25.
- Aizawa, K. (2010). The coupling-constitution fallacy revisited. *Cognitive Systems Research*, 11:332–342.
- Aizawa, K. (2015). What is this cognition that is supposed to be embodied? *Philosophical Psychology*, 28(6):755–775.
- Alday, P. M. (2019). How much baseline correction do we need in erp research? extended glm model can replace baseline correction while lifting its limits. *Psychophysiology*, 56(12):e13451.

- Alexander, W. H. and Brown, J. W. (2010). Computational models of performance monitoring and cognitive control. *Topics in Cognitive Science*, 2(4):658–677.
- Allard, E., Canzoneri, E., Adler, D., Morélot-Panzini, C., Bello-Ruiz, J., Herbelin, B., Blanke, O., and Similowski, T. (2017). Interferences between breathing, experimental dyspnoea and bodily self-consciousness. *Scientific Reports*, 7(1):9990.
- Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., and Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *eLife*, 5.
- Allen, M. and Friston, K. J. (2018). From cognitivism to autopoiesis: Towards a computational framework for the embodied mind. *Synthese*, 195(6):2459–2482.
- Allen, M., Levy, A., Parr, T., and Friston, K. J. (2019). In the body’s eye: The computational anatomy of interoceptive inference. *bioRxiv*.
- Allen, M. and Tsakiris, M. (2018). The body as first prior: Interoceptive predictive processing and the primacy of self-models. In Tsakiris, M. and De Preester, H., editors, *The interoceptive mind: From homeostasis to awareness*, pages 27–45. Oxford: Oxford University Press.
- Alsmith, A. J. T. and de Vignemont, F. (2012). Embodying the mind and representing the body. *Review of Philosophy & Psychology*, 3:1–13.
- Altmann, G. T. M. and Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33(4):583–609.
- Andrews, M. (2021). The math is not the territory: Navigating the free energy principle. *Biology & Philosophy*, 36(30).
- Andrillon, T., Burns, A., MacKay, T., Windt, J. M., and Tsuchiya, N. (2021a). Predicting lapses of attention with sleep-like slow waves. *Nature Communications*, 12(3657):1–12.
- Andrillon, T., Burns, A., MacKay, T., Windt, J. M., and Tsuchiya, N. (2021b). Predicting lapses of attention with sleep-like slow waves. Online. <https://osf.io/ey3ca>.
- Andrillon, T., Windt, J. M., Silk, T., Drummond, S. P. A., Bellgrove, M. A., and Tsuchiya, N. (2019). Does the mind wander when the brain takes a break? local sleep in wakefulness, attentional lapses and mind-wandering. *Frontiers in Neuroscience*, 13:949.
- Antrobus, J. S. (1968). Information theory and stimulus-independent thought. *British Journal of Psychology*, 59(4):423–430.

- Antrobus, J. S., Singer, J. L., and Greenberg, S. (1966). Studies in the stream of consciousness: Experimental enhancement and suppression of spontaneous cognitive processes. *Perceptual & Motor Skills*, 23:399–417.
- Apps, M. A. J. and Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, 41:85–97.
- Arnal, L. H. and Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, 16(7):390–8.
- Aspell, J. E., Heydrich, L., Marillier, G., Lavanchy, T., Herbelin, B., and Blanke, O. (2013). Turning body and self inside out: Visualized heartbeats alter bodily self-consciousness and tactile perception. *Psychological Science*, 24(12):2445–2453.
- Aspell, J. E., Lenggenhager, B., and Blanke, O. (2012). Multisensory perception and bodily self-consciousness: From out-of-body to inside-body experience. In Murray, M. M. and Wallace, M. T., editors, *The neural bases of multisensory processes*, chapter 24. Boca Raton, FL.: CRC Press.
- Atzil, S., Gao, W., Fradkin, I., and Barrett, L. F. (2018). Growing a social brain. *Nature Human Behaviour*, 2(9):624–636.
- Ax, A. F. (1964). Goals and methods of psychophysiology. *Psychophysiology*, 1:8–25.
- Azevedo, R. T., Garfinkel, S. N. and Critchley, H. D., and Tsakiris, M. (2017). Cardiac afferent activity modulates the expression of racial stereotypes. *Nature Communications*, 8:13854.
- Azzalini, D., Rebollo, I., and Tallon-Baudry, C. (2019). Visceral signals shape brain dynamics and cognition. *Trends in Cognitive Sciences*, 23(6):488–509.
- Baayen, H. R., Vasishth, S., Kliegl, R., and Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory & Language*, 94:206–234.
- Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., and Friston, K. J. (2017). The depressed brain: An evolutionary systems theory. *Trends in Cognitive Sciences*, 21(3):182–194.
- Baltieri, M., Buckley, C. L., and Bruineberg, J. (2020). Predictions in the eye of the beholder: An active inference account of wett governors. In Bongard, J., Lovato, J., Hebert-Dufresne, L., Dasari, R., and Soros, L., editors, *ALIFE 2020: The 2020 conference on artificial life*, volume 32, pages 121–129. MIT Press.

- Baltzell, L. S., Srinivasan, R., and Richards, V. M. (2017). The effect of prior knowledge and intelligibility on the cortical entrainment response to speech. *Journal of Neurophysiology*, 118(6):3144–3151.
- Bard, P. A. (1928). A diencephalic mechanism for the expression of rage with special reference to the central nervous system. *American Journal of Physiology*, 84:490–513.
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory & Language*, 68(3).
- Barrett, L. (2011). *Beyond the brain: How body and environment shape animal and human minds*. Princeton & Oxford: Princeton University Press.
- Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive & Affective Neuroscience*, 12(1):1–23.
- Barrett, L. F. and Bar, M. (2009). See it with feeling: Affective predictions during object perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521):1325–1334.
- Barrett, L. F., Quigley, K. S., and Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B*, 371(20160011):1–17.
- Barrett, L. F. and Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, 16(7):419–429.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences*, 22(4):577–609.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59:617–645.
- Bastiaansen, M., Magyari, L., and Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience*, 22(7):1333–1347.
- Bastin, J., Deman, P., David, O., Gueguen, M., Benis, D., Minotti, L., Hoffman, D., Combrisson, E., Kujala, J., Perrone-Bertolotti, M., Kahane, P., Lachaux, J.-P., and Jerbi, K. (2017). Direct recordings from human anterior insula reveal its leading role within the error-monitoring network. *Cerebral Cortex*, 27(2):1545–1557.
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, 76(4):695–711.

- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.
- Bell, A. J. and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159.
- Bellgrove, M. A., Hawi, Z., Gill, M., and Robertson, I. H. (2006). The cognitive genetics of attention deficit hyperactivity disorder (adhd): Sustained attention as a candidate phenotype. *Cortex*, 42(6):838–845.
- Bellgrove, M. A., Hawi, Z., Kirley, A., Gill, M., and Robertson, I. H. (2005). Dissecting the attention deficit hyperactivity disorder (adhd) phenotype: sustained attention, response variability and spatial attentional asymmetries in relation to dopamine transporter (dat1) genotype. *Neuropsychologia*, 43(13):1847–1857.
- Benarroch, E. E. (1993). The central autonomic network: Functional organization, dysfunction, and perspective. *Mayo Clinic Proceedings*, 68(10):988–1001.
- Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. New York, Toronto, London: McGraw-Hill Book Company.
- Berntson, G. G. and Khalsa, S. S. (2021). Neural circuits of interoception. *Trends in Neurosciences*, 44(1):17–28.
- Betka, S., Canzoneri, E., Adler, D., Herbelin, B., Bello-Ruiz, J., Kannape, O. A., Similowski, T., and Blanke, O. (2020). Mechanisms of the breathing contribution to bodily self-consciousness in healthy humans: Lessons from machine-assisted breathing? *Psychophysiology*, 57(8):e13564.
- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., and Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, 7(3):295–301.
- Birren, J. E., Cardon, Jr., P. V., and Phillips, S. L. (1963). Reaction time as a function of the cardiac cycle in young adults. *Science*, 140(3563):195–196.
- Blanchard, N., Bixler, R., Joyce, T., and D’Mello, S. K. (2014). Automated physiological-based detection of mind wandering during learning. In Trausan-Matu, S., Boyer, K. E., Crosby, M., and Panourgia, K., editors, *ITS 2014: Intelligent Tutoring Systems*, volume 8474 of *Lecture Notes in Computer Science*, pages 55–60.
- Blank, H. and Davis, M. H. (2016). Prediction errors but not sharpened signals simulate multivoxel fmri patterns during speech perception. *PLoS Biology*, 14(11):e1002577.

- Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nature Reviews Neuroscience*, 13(8):556–571.
- Blanke, O. and Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences*, 13(1):7–13.
- Blanke, O., Slater, M., and Serino, A. (2015). Behavioral, neural, and computational principles of bodily self-consciousness. *Neuron*, 88(1):145–166.
- Blankenship, A. G. and Feller, M. B. (2010). Mechanisms underlying spontaneous patterned activity in developing neural circuits. *Nature Reviews Neuroscience*, 11(1):18–29.
- Block, N. (2005). Review of alva noë: Action in perception. *Journal of Philosophy*, 102(5):259–272.
- Boersma, P. and Weenink, D. (2011). Praat: Doing phonetics by computer. Online.
- Bohlin, G. and Kjellberg, A. (1979). Orienting activity in two stimulus paradigms as reflected in heart rate. In Kimmel, H. D., van Olst, E. H., and Orlebeke, J. F., editors, *The orienting reflex in humans*, pages 169–198. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Bonaz, B., Lane, R. D., Oshinsky, M. L., Kenny, P. J., Sinha, R., Mayer, E. A., and Critchley, H. D. (2021). Diseases, disorders, and comorbidities of interoception. *Trends in Neurosciences*, 44(1):39–51.
- Bonhage, C. E., Meyer, L., Gruber, T., Friederici, A. D., and Mueller, J. L. (2017). Oscillatory eeg dynamics underlying automatic chunking during sentence processing. *NeuroImage*, 152:647–657.
- Bonnefond, A., Doignon-Camus, N., Touzalin-Chretien, P., and Dufour, A. (2010). Vigilance and intrinsic maintenance of alert state: An erp study. *Behavioural Brain Research*, 211(2):185–190.
- Bornkessel-Schlesewsky, I. and Schlewsky, M. (2013). Reconciling time, space and function: A new dorsal-ventral stream model of sentence comprehension. *Brain & Language*, 125(1):60–76.
- Bornkessel-Schlesewsky, I. and Schlewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*, 10:298.
- Bornkessel-Schlesewsky, I., Schlewsky, M., Small, S. L., and Rauschecker, J. P. (2015). Neurobiological roots of language in primate audition: Common computational properties. *Trends in Cognitive Sciences*, 19(3):142–150.

- Botvinick, M. and Cohen, J. (1998). Rubber hands ‘feel’ touch that eyes see. *Nature*, 391(6669):756.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., and Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3):624–652.
- Boullin, J. and Morgan, J. M. (2005). The development of cardiac rhythm. *Heart*, 91(7):874–875.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4):433–436.
- Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L., and Snyder, A. (2001). Anterior cingulate cortex and response conflict: Effects of frequency, inhibition and errors. *Cerebral Cortex*, 11(9):825–836.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Brodbeck, C. and Simon, J. Z. (2020). Continuous speech processing. *Current Opinion in Physiology*, 18:25–31.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47(1-3):139–159.
- Brown, G. L. (1927). Daydreams: A cause of mind wandering and inferior scholarship. *Journal of Educational Research*, 15(4):276–279.
- Brown, H., Friston, K. J., and Bestmann, S. (2011). Active inference, attention, and motor preparation. *Frontiers in Psychology*, 2(218).
- Bruineberg, J., Kiverstein, J., and Rietveld, E. (2018). The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. *Synthese*, 195(6):2417–2444.
- Bruineberg, J. and Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience*, 8(599).
- Bury, G., García-Huésca, M., Bhattacharya, J., and Ruiz, M.-H. (2019). Cardiac afferent activity modulates early neural signature of error detection during skilled performance. *NeuroImage*, 199:704–717.
- Buzsáki, G. (2002). Theta oscillations in the hippocampus. *Neuron*, 33(3):325–340.
- Buzsáki, G. (2010). Neural syntax: Cell assemblies, synapse ensembles, and readers. *Neuron*, 68(3):362–385.

- Buzsáki, G. and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679):1926–1929.
- Buzsáki, G. and Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience*, 16(2):130–138.
- Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G. (2007). Psychophysiological science: Interdisciplinary approaches to classic questions about the mind. In Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G., editors, *Handbook of psychophysiology*, chapter 1, pages 1–16. Cambridge: Cambridge University Press, 3rd edition.
- Callaway, 3rd, E. and Layne, R. S. (1964). Interaction between the visual evoked response and two spontaneous biological rhythms: The eeg alpha cycle and the cardiac arousal cycle. *Annals of the New York Academy of Sciences*, 112:421–431.
- Cameron, O. G. (2002). *Visceral sensory neuroscience: Interoception*. Oxford: Oxford University Press.
- Cannon, W. B. (1927). The james-lange theory of emotions: A critical examination and an alternative theory. by walter b. cannon, 1927. *American Journal of Psychology*, 39:106–124.
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3):200–219.
- Chan, R. C. K., Wang, Y., Cheung, E. F. C., Cui, J., Deng, Y., Yuan, Y., Ma, Z., Yu, X., Li, Z., and Gong, Q. (2009). Sustained attention deficit along the psychosis proneness continuum: A study on the sustained attention to response task (sart). *Cognitive & Behavioral Neurology*, 22(3):180–185.
- Cheetham, M., Cepeda, C., and Gamboa, H. (2016). Automated detection of mind wandering: A mobile application. In Gilbert, J., Azhari, H., Ali, H., Quintao, C., Sliwa, J., Ruiz, C., Fred, A., and Gamboa, H., editors, *Proceedings of the 9th international joint conference on biomedical engineering systems and technologies*, volume 4, pages 198–205. SCITEPRESS.
- Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge, MA: MIT Press.
- Chemero, A. (2013). Radical embodied cognitive science. *Review of General Psychology*, 17(2):145–150.
- Chen, W. G., Schloesser, D., Arensdorf, A. M., Simmons, J. M., Cui, C., Valentino, R., Gnadt, J. W., Nielsen, L., Hillaire-Clarke, C. S., Spruance, V., Horowitz, T. S., Vallejo, Y. F., and Langevin, H. M. (2021). The emerging science of interoception:

- Sensing, integrating, interpreting, and regulating signals within the self. *Trends in Neurosciences*, 44(1):3–16.
- Cheng, G., Zhou, X., Qu, J., Ashwell, K. W. S., and Paxinos, G. (2004). Central vagal sensory and motor connections: Human embryonic and fetal development. *Autonomic Neuroscience: Basic & Clinical*, 114(1–2):83–96.
- Cheng, G., Zhu, H., Zhou, X., Qu, J., Ashwell, K. W. S., and Paxinos, G. (2006). Development of the human nucleus of the solitary tract: A cyto- and chemoarchitectural study. *Autonomic Neuroscience: Basic & Clinical*, 128(1–2):76–95.
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 25(5):975–979.
- Cheyne, J. A., Carriere, J. S. A., and Smilek, D. (2006). Absent-mindedness: Lapses of conscious awareness and everyday cognitive failures. *Consciousness & Cognition*, 15(3):578–592.
- Cheyne, J. A., Carriere, J. S. A., Solman, G. J. F., and Smilek, D. (2011). Challenge and error: Critical events and attention-related errors. *Cognition*, 121(3):437–446.
- Cheyne, J. A., Solman, G. J. F., Carriere, J. S. A., and Smilek, D. (2009). Anatomy of an error: A bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition*, 111(1):98–113.
- Christensen, R. H. B. (2019). ordinal – regression models for ordinal data.
- Christiansen, M. H. and Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral Brain Sciences*, 39(e62).
- Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., and Schooler, J. W. (2009). Experience sampling during fmri reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences of the United States of America*, 106(21):8719–8724.
- Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N., and Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: A dynamic framework. *Nature Reviews Neuroscience*, 17(11):718–731.
- Ciaunica, A., Constant, A., Preissl, H., and Fotopoulou, A. (2021). The first prior: From co-embodiment to co-homeostasis in early life. *Preprint*.
- Ciaunica, A. and Crucianelli, L. (2019). Minimal self-awareness: From within a developmental perspective. *Journal of Consciousness Studies*, 26(3–4):207–226.

- Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.
- Clark, A. (2008a). Pressing the flesh: A tension in the study of the embodied, embedded mind? *Philosophy & Phenomenological Research*, 76(1):37–59.
- Clark, A. (2008b). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford: Oxford University Press.
- Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral Brain Sciences*, 36(3):181–253.
- Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford: Oxford University Press.
- Clark, A. (2017a). Busting out: Predictive brains, embodied minds, and the puzzle of the evidentiary veil. *Noûs*, 51(4):727–753.
- Clark, A. (2017b). How to knit your own markov blanket: Resisting the second law with metamorphic minds. In Metzinger, T. and Wiese, W., editors, *Philosophy and Predictive Processing*, chapter 3, pages 1–19. Frankfurt am Main: MIND Group.
- Clark, A. (2018). A nice surprise? predictive processing and the active pursuit of novelty. *Phenomenology & the Cognitive Sciences*, 17(3):521–534.
- Clark, J. E., Watson, S., and Friston, K. J. (2018). What is mood? a computational perspective. *Psychological Medicine*, 48(14):2277–2284.
- Clos, M., Langner, R., Meyer, M., Oechslin, M. S., Zilles, K., and Eickhoff, S. B. (2014). Effects of prior information on decoding degraded speech: An fmri study. *Human Brain Mapping*, 35(1):61–74.
- Coleman, W. M. (1921). The psychological significance of bodily rhythms. *Journal of Comparative & Physiological Psychology*, 1:213–220.
- Coles, M. G. H. and Duncan-Johnson, C. C. (1975). Cardiac activity and information processing: The effects of stimulus significance, and detection and response requirements. *Journal of Experimental Psychology: Human Perception & Performance*, 1(4):418–428.
- Coles, M. G. H. and Strayer, D. L. (1985). The psychophysiology of the cardiac cycle time effect. In Orlebeke, J. F., Mulder, G., and van Doornen, L. J. P., editors, *Psychophysiology of cardiovascular control: Models, methods, and data*, pages 517–548. New York & London: Plenum Press.

- Colombetti, G. (2014). *The feeling body: Affective science meets the enactive mind*. Cambridge, MA: MIT Press.
- Colombo, M. and Wright, C. (2018). First principles in the life sciences: The free-energy principle, organicism, and mechanism. *Synthese*.
- Connor, W. H. and Lang, P. J. (1969). Cortical slow-wave and cardiac rate responses in stimulus orientation and reaction time conditions. *Journal of Experimental Psychology*, 82(2):310–320.
- Corcoran, A. W. (2019). Cephalopod molluscs, causal models, and curious minds. *Animal Sentience*, 4(26):13.
- Corcoran, A. W., Macefield, V. G., and Hohwy, J. (2021). Be still my heart: Cardiac regulation as a mode of uncertainty reduction. *Psychonomic Bulletin & Review*, 28(4):1211–1223.
- Corcoran, A. W., Pezzulo, G., and Hohwy, J. (2018). Commentary: Respiration-entrained brain rhythms are global but often overlooked. *Frontiers in Systems Neuroscience*, 12:25.
- Craig, A. D. (2002). How do you feel? interoception: The sense of the physiological condition of the body. *Nature Reviews Neuroscience*, 3:655–666.
- Critchley, H. D. and Garfinkel, S. N. (2017). Interoception and emotion. *Current Opinion in Psychology*, 17:7–14.
- Critchley, H. D. and Garfinkel, S. N. (2018). The influence of physiological signals on cognition. *Current Opinion in Behavioral Sciences*, 19:13–18.
- Crone, E. A., Bunge, S. A., de Klerk, P., and van der Molen, M. W. (2005). Cardiac concomitants of performance monitoring: context dependence and individual differences. *Brain Research: Cognitive Brain Research*, 23(1):93–106.
- Crone, E. A., Somsen, R. J. M., van Beek, B., and van der Molen, M. W. (2004). Heart rate and skin conductance analysis of antecedents and consequences of decision making. *Psychophysiology*, 41(4):531–540.
- Crone, E. A., van der Veen, F. M., van der Molen, M. W., Somsen, R. J. M., van Beek, B., and Jennings, J. R. (2003). Cardiac concomitants of feedback processing. *Biological Psychology*, 64(1-2):143–156.
- Cross, Z. R., Corcoran, A. W., Schlesewsky, M., Kohler, M. J., and Bornkessel-Schlesewsky, I. (2020). Oscillatory and aperiodic neural activity jointly predict grammar learning. *bioRxiv*.

- Cross, Z. R., Kohler, M. J., Schlesewsky, M., Gaskell, M. G., and Bornkessel-Schlesewsky, I. (2018). Sleep-dependent memory consolidation and incremental sentence comprehension: Computational dependencies during language learning as revealed by neuronal oscillations. *Frontiers in Human Neuroscience*, 12:18.
- Crosse, M. J., Di Liberto, G. M., Bednar, A., and Lalor, E. C. (2016). The multivariate temporal response function (mtrf) toolbox: A matlab toolbox for relating neural signals to continuous stimuli. *Frontiers in Human Neuroscience*, 10:604.
- Damasio, A. (1994). *Descartes' error: Emotion, reason and the human brain*. New York, NY: G. P. Putnam's Sons.
- Damasio, A. (2010). *Self comes to mind: Constructing the conscious brain*. New York, NY: Random House.
- Damasio, A. (2018). *The strange order of things: Life, feeling, and the making of cultures*. New York, NY: Pantheon Books.
- Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 351(1346):1413–1420.
- Danev, S. G. and de Winter, C. R. (1971). Heart rate deceleration after erroneous responses: A phenomenon complicating the use of heart rate variability for assessing mental load. *Psychologische Forschung*, 35(1):27–34.
- Dang, J. S., Figueroa, I. J., and Helton, W. S. (2018). You are measuring the decision to be fast, not inattention: The sustained attention to response task does not measure sustained attention. *Experimental Brain Research*, 236(8):2255–2262.
- Davis, M. H. and Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, 23(8):3423–3431.
- Davis, M. H. and Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, 229(1-2):132–147.
- Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, 134(2):222–241.
- De Pascalis, V., Barry, R. J., and Sparita, A. (1995). Decelerative changes in heart rate during recognition of visual stimuli: Effects of psychological stress. *International Journal of Psychophysiology*, 20(1):21–31.

- de Vignemont, F. (2011). Embodiment, ownership and disownership. *Consciousness & Cognition*, 20(1):82–93.
- Deane, G. (2020). Dissolving the self: Active inference, psychedelics, and ego-dissolution. *Philosophy & the Mind Sciences*, 1(1):1–27.
- Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., and Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage*, 24(1):21–33.
- Dell, G. S. and Chang, F. (2014). The p-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634):20120394.
- DeLong, K. A., Troyer, M., and Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language & Linguistics Compass*, 8(12):631–645.
- Delorme, A. and Makeig, S. (2004). Eeglab: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.
- Desmet, C., Fias, W., and Brass, M. (2011). Performance monitoring at the task and the response level. *Reviews in the Neurosciences*, 22(5):575–581.
- Di Liberto, G. M., Crosse, M. J., and Lalor, E. C. (2018). Cortical measures of phoneme-level speech encoding correlate with the perceived clarity of natural speech. *eNeuro*, 5(2).
- Di Lullo, E. and Kriegstein, A. R. (2017). The use of brain organoids to investigate neural development and disease. *Nature Reviews Neuroscience*, 18(10):573–584.
- Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology & the Cognitive Sciences*, 4(4):429–452.
- Di Paolo, E. A. (2018). The enactive conception of life. In Newen, A., de Bruin, L., and Gallagher, S., editors, *The Oxford handbook of 4E cognition*, chapter 4, pages 71–94. Oxford: Oxford University Press.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64:135–168.
- Dillard, M. B., Warm, J. S., Funke, G. J., Funke, M. E., Finomore, Jr, V. S., Matthews, G., Shaw, T. H., and Parasuraman, R. (2014). The sustained attention to response task (sart) does not promote mindlessness during vigilance performance. *Human Factors*, 56(8):1364–1379.

- Dimitrijevic, A., Smith, M. L., Kadis, D. S., and Moore, D. R. (2019). Neural indices of listening effort in noisy environments. *Scientific Reports*, 9(1):11278.
- Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, 19(1):158–164.
- Ding, N. and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, 109(29):11854–11859.
- Ding, N. and Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, 8:311.
- DiPietro, J. A., Costigan, K. A., and Voegtline, K. M. (2015). Studies in fetal behavior: Revisited, renewed, and reimagined. *Monographs of the Society for Research in Child Development*, 80(3):vii–94.
- Dockree, P. M., Kelly, S. P., Roche, R. A. P., Hogan, M. J., Reilly, R. B., and Robertson, I. H. (2004). Behavioural and physiological impairments of sustained attention after traumatic brain injury. *Brain Research: Cognitive Brain Research*, 20(3):403–414.
- Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, 85 Pt 2:761–768.
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, 30:412–431.
- Duncan-Johnson, C. C. and Coles, M. G. H. (1974). Heart rate and disjunctive reaction time: The effects of discrimination requirements. *Journal of Experimental Psychology*, 103(6):1160–1168.
- Edwards, L., Inui, K., Ring, C., Wang, X., and Kakigi, R. (2008). Pain-related evoked potentials are modulated across the cardiac cycle. *Pain*, 137(3):488–494.
- Edwards, L., Ring, C., McIntyre, D., Carroll, D., and Martin, U. (2007). Psychomotor speed in hypertension: Effects of reaction time components, stimulus modality, and phase of the cardiac cycle. *Psychophysiology*, 44(3):459–468.
- Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science*, 317(5841):1048.
- Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., and Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience*, 30(21):7179–7186.

- Elliott, R. (1969). Tonic heart rate: Experiments on the effects of collative variables lead to a hypothesis about its motivational significance. *Journal of Personality & Social Psychology*, 12(3):211–228.
- Elliott, R., Bankart, B., and Light, T. (1970). Differences in the motivational significance of heart rate and palmar conductance: Two tests of a hypothesis. *Journal of Personality & Social Psychology*, 14(2):166–172.
- Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, 2(10):704–716.
- Faber, M., Bixler, R., and D’Mello, S. K. (2018). An automated behavioral measure of mind wandering during computerized reading. *Behavior Research Methods*, 50(1):134–150.
- Fabry, R. E. (2017a). Betwixt and between: The enculturated predictive processing approach to cognition. *Synthese*.
- Fabry, R. E. (2017b). Predictive processing and cognitive development. In Metzinger, T. and Wiese, W., editors, *Philosophy and Predictive Processing*, chapter 13, pages 1–18. Frankfurt am Main: MIND Group.
- Farrin, L., Hull, L., Unwin, C., Wykes, T., and David, A. (2003). Effects of depressed mood on objective and subjective measures of attention. *Journal of Neuropsychiatry & Clinical Neurosciences*, 15(1):98–104.
- Fasiolo, M., Nedellec, R., Goude, Y., and Wood, S. N. (2018). Scalable visualisation methods for modern generalized additive models. *arXiv*, 1809.10632.
- Favela, L. H. (2014). Radical embodied cognitive neuroscience: Addressing ”grand challenges” of the mind sciences. *Front. Hum. Neurosci.*, 8(796):1–10.
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4):491–505.
- Feldman, H. and Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4(215):1–23.
- Feller, M. B. (1999). Spontaneous correlated activity in developing neural circuits. *Neuron*, 22(4):653–656.
- Fernandez Velasco, P. and Loev, S. (2020). Affective experience in the predictive mind: A review and new integrative account. *Synthese*.

- Fiacconi, C. M., Peter, E. L., Owais, S., and Köhler, S. (2016). Knowing by heart: Visceral feedback shapes recognition memory judgments. *Journal of Experimental Psychology: General*, 145(5):559–572.
- Fiehler, K., Ullsperger, M., Grigutsch, M., and von Cramon, D. Y. (2004). Cardiac responses to error processing and response conflict. In Ullsperger, M. and Falkenstein, M., editors, *Errors, conflicts, and the brain. Current opinions on performance monitoring*, pages 135–140. Leipzig: MPI for Human Cognitive & Brain Sciences.
- Flexman, J. E. (1974). Respiratory phase and visual signal detection. *Perception & Psychophysics*, 16(2):337–339.
- Fodor, J. A. (1975). *The language of thought*. Language and thought series. Cambridge, MA: Harvard University Press.
- Fodor, J. A. (1981). *Representations: Philosophical essays on the foundations of cognitive science*. Brighton: Harvester Press.
- Foglia, L. and Wilson, R. A. (2013). Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(3):319–325.
- Fotopoulou, A. and Tsakiris, M. (2017). Mentalizing homeostasis: The social origins of interoceptive inference. *Neuropsychanalysis*, pages 1–26.
- Foulsham, T., Farley, J., and Kingstone, A. (2013). Mind wandering in sentence reading: Decoupling the link between mind and eye. *Canadian Journal of Experimental Psychology*, 67(1):51–59.
- Fox, J. and Weisberg, S. (2019). *An R companion to applied regression*. Thousand Oaks CA: Sage, 3rd edition.
- Foxe, J. J., Morie, K. P., Laud, P. J., Rowson, M. J., de Bruin, E. A., and Kelly, S. P. (2012). Assessing the effects of caffeine and theanine on the maintenance of vigilance during a sustained attention task. *Neuropharmacology*, 62(7):2320–2327.
- Fraga González, G., Smit, D. J. A., van der Molen, M. J. W., Tijms, J., de Geus, E. J. C., and van der Molen, M. W. (2019). Probability learning and feedback processing in dyslexia: A performance and heart rate analysis. *Psychophysiology*, 56(12):e13460.
- Friauf, E. and Lohmann, C. (1999). Development of auditory brainstem circuitry: Activity-dependent and activity-independent processes. *Cell & Tissue Research*, 297(2):187–195.
- Fridman, J., Barrett, L. F., Wormwood, J. B., and Quigley, K. S. (2019). Applying the theory of constructed emotion to police decision making. *Frontiers in Psychology*, 10:1946.

- Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9(10):474–480.
- Fries, P. (2015). Rhythms for cognition: Communication through coherence. *Neuron*, 88(1):220–235.
- Friston, K. J. (2002). Functional integration and inference in the brain. *Progress in Neurobiology*, 68(2):113–143.
- Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B*, 360(1456):815–836.
- Friston, K. J. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, 13(7):293–301.
- Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138.
- Friston, K. J. (2013). Life as we know it. *Journal of the Royal Society Interface*, 10(86):20130475.
- Friston, K. J. (2018). Does predictive coding have a future? *Nature Neuroscience*, 21(8):1019–1021.
- Friston, K. J. (2019a). A free energy principle for a particular physics. *arXiv*, page 1906.10184.
- Friston, K. J. (2019b). Waves of prediction. *PLoS Biology*, 17(10):e3000426.
- Friston, K. J., Adams, R. A., Perrinet, L., and Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, 3(151):1–20.
- Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O’Doherty, J., and Pezzulo, G. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews*, 68:862–879.
- Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., and Pezzulo, G. (2017a). Active inference: A process theory. *Neural Computation*, 29(1):1–49.
- Friston, K. J. and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B*, 364(1521):1211–1221.
- Friston, K. J., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, 100(1-3):70–87.

- Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., and Ondobaka, S. (2017b). Active inference, curiosity and insight. *Neural Computation*, 29(10):2633–2683.
- Friston, K. J., Mattout, J., and Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, 104(1-2):137–160.
- Friston, K. J., Rigoli, F., Ognibene, D., Mathys, C. D., Fitzgerald, T., and Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, 6(4):187–224.
- Friston, K. J., Sajid, N., Quiroga-Martinez, D. R., Parr, T., Price, C. J., and Holmes, E. (2021). Active listening. *Hearing Research*, 399:107998.
- Friston, K. J. and Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, 159(3):417–458.
- Friston, K. J., Stephan, K. E., Montague, R., and Dolan, R. J. (2014). Computational psychiatry: The brain as a phantastic organ. *Lancet Psychiatry*, 1(2):148–158.
- Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, 4(1):14–21.
- Gallagher, S. (2005). *How the body shapes the mind*. Oxford: Oxford University Press.
- Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*. Oxford: Oxford University Press.
- Gallese, V. and Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in Cognitive Sciences*, 15(11):512–519.
- Gallistel, C. R. (1989). Animal cognition: The representation of space, time and number. *Annual Review of Psychology*, 40:155–189.
- Galvez-Pol, A., McConnell, R., and Kilner, J. M. (2020). Active sampling in visual search is coupled to the cardiac cycle. *Cognition*, 196:104149.
- Garfinkel, S. N. and Critchley, H. D. (2016). Threat and the body: How the heart supports fear processing. *Trends in Cognitive Sciences*, 20(1):34–46.
- Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., and Critchley, H. D. (2015). Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biological Psychology*, 104:65–74.
- Germana, J. (1969). Central efferent processes and autonomic-behavioral integration. *Psychophysiology*, 6(1):78–90.

- Gerrans, P. and Murray, R. J. (2020). Interoceptive active inference and self-representation in social anxiety disorder (sad): Exploring the neurocognitive traits of the sad self. *Neuroscience of Consciousness*, page niaa026.
- Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, 2:130.
- Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum. *Frontiers in Psychology*, 3:238.
- Ghitza, O. (2013). The theta-syllable: A unit of speech information defined by cortical function. *Frontiers in Psychology*, 4:138.
- Ghitza, O. and Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, 66(1-2):113–126.
- Giambra, L. M. (1995). A laboratory method for investigating influences on switching attention to task-unrelated imagery and thought. *Consciousness & Cognition*, 4(1):1–21.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA.: Houghton Mifflin.
- Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C., and Kleinschmidt, A. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral Cortex*, 14(3):247–255.
- Giraud, A.-L. and Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4):511–517.
- Godfrey-Smith, P. (1996). *Complexity and the function of mind in nature*. Cambridge Studies in Philosophy and Biology. Cambridge: Cambridge University Press.
- Goldman, A. and de Vignemont, F. (2009). Is social cognition embodied? *Trends in Cognitive Sciences*, 13(4):154–159.
- Goldman, A. I. (2012). A moderate approach to embodied cognitive science. *Review of Philosophy & Psychology*, 3(1):71–88.

- Goldman, A. I. (2016). Reply to firestone. In McLaughlin, B. P. and Kornblith, H., editors, *Goldman and his critics*, chapter 15, pages 335–336. West Sussex: John Wiley & Sons, Inc.
- Goodman, C. S. and Shatz, C. J. (1993). Developmental mechanisms that generate precise patterns of neuronal connectivity. *Cell*, 72 Suppl:77–98.
- Graham, F. K. (1979). Distinguishing among orienting, defense, and startle reflexes. In Kimmel, H. D., van Olst, E. H., and Orlebeke, J. F., editors, *The orienting reflex in humans*, chapter 8, pages 137–167. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Graham, F. K. and Clifton, R. K. (1966). Heart-rate change as a component of the orienting response. *Psychological Bulletin*, 65(5):306–320.
- Green, D. M. and Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.
- Groen, Y., Wijers, A. A., Mulder, L. J. M., Minderaa, R. B., and Althaus, M. (2007). Physiological correlates of learning by performance feedback in children: a study of eeg event-related potentials and evoked heart rate. *Biological Psychology*, 76(3):174–187.
- Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., and Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, 11(12):e1001752.
- Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, 31:423–445.
- Grossberg, S. and Kazerounian, S. (2011). Laminar cortical dynamics of conscious speech perception: Neural model of phonemic restoration using subsequent context in noise. *Journal of the Acoustical Society of America*, 130(1):440–460.
- Grund, M., Al, E., Pabst, M., Dabbagh, A., Stephani, T., Nierhaus, T., and Villringer, A. (2021). Respiration, heartbeat, and conscious tactile perception. *bioRxiv*.
- Gu, X., FitzGerald, T. H. B., and Friston, K. J. (2019). Modeling subjective belief states in computational psychiatry: Interoceptive inference as a candidate framework. *Psychopharmacology*, 236(8):2405–2412.
- Gu, X., Hof, P. R., Friston, K. J., and Fan, J. (2013). Anterior insular cortex and emotional awareness. *Journal of Comparative Neurology*, 521(15):3371–3388.
- Guediche, S., Blumstein, S. E., Fiez, J. A., and Holt, L. L. (2014). Speech perception under adverse conditions: Insights from behavioral, computational, and neuroscience research. *Frontiers in Systems Neuroscience*, 7:126.

- Haegens, S. and Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience & Biobehavioral Reviews*, 86:150–165.
- Hagoort, P. (2017). The core and beyond in the language-ready brain. *Neuroscience & Biobehavioral Reviews*, 81(Pt B):194–204.
- Hagoort, P. (2019). The neurobiology of language beyond single-word processing. *Science*, 366(6461):55–58.
- Hagoort, P. and van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481):801–811.
- Hajcak, G., McDonald, N., and Simons, R. F. (2003). To err is autonomic: Error-related brain potentials, ans activity, and post-error compensatory behavior. *Psychophysiology*, 40(6):895–903.
- Haken, H. (1983). *Synergetics, an introduction: Nonequilibrium phase transitions and self-organization in physics, chemistry, and biology*. New York, NY: Springer-Verlag.
- Halle, M. and Stevens, K. N. (1959). Analysis by synthesis. In Wathen-Dunn, W. and Woods, L. E., editors, *Proceedings of the seminar on speech comprehension and processing*, volume 2. Bedford, MA: USAF Cambridge Research Center.
- Hanganu-Opatz, I. L. (2010). Between molecules and experience: Role of early patterns of coordinated activity for the development of cortical maps and sensory abilities. *Brain Research Reviews*, 64(1):160–176.
- Hanslmayr, S., Staresina, B. P., and Bowman, H. (2016). Oscillations and episodic memory: Addressing the synchronization/desynchronization conundrum. *Trends in Neurosciences*, 39(1):16–25.
- Haugeland, J. (1998). Mind embodied and embedded. In Haugeland, J., editor, *Having thought: Essays in the metaphysics of mind*, pages 207–237. Cambridge, MA: Harvard University Press.
- Hauswald, A., Keitel, A., Chen, Y.-P., Rösch, S., and Weisz, N. (2020). Degradation levels of continuous speech affect neural speech tracking and alpha power differently. *European Journal of Neuroscience*.
- Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behavior Reseach Methods, Instruments, & Computers*, 27(1):46–51.
- He, J., Becic, E., Lee, Y.-C., and McCarley, J. S. (2011). Mind wandering behind the wheel: Performance and oculomotor correlates. *Human Factors*, 53(1):13–21.

- Head, J. and Helton, W. S. (2013). Perceptual decoupling or motor decoupling? *Consciousness & Cognition*, 22(3):913–919.
- Heald, S. L. M. and Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, 8:35.
- Heilbron, M. and Chait, M. (2018). Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience*, 389:54–73.
- Helton, W. S. (2009). Impulsive responding and the sustained attention to response task. *Journal of Clinical & Experimental Neuropsychology*, 31(1):39–47.
- Herman, A. M., Esposito, G., and Tsakiris, M. (2021). Body in the face of uncertainty: The role of autonomic arousal and interoception in decision-making under risk and ambiguity. *Psychophysiology*, 58(8):e13840.
- Herman, A. M. and Tsakiris, M. (2020). Feeling in control: The role of cardiac timing in the sense of agency. *Affective Science*, 1:155–171.
- Hervais-Adelman, A., Carlyon, R. P., Johnsrude, I. S., and Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Language & Cognitive Processes*, 27(7/8):1145–1166.
- Hervé, M. (2021). Rvaidememoire: Testing and plotting procedures for biostatistics.
- Hesp, C., Smith, R., Parr, T., Allen, M., Friston, K. J., and Ramstead, M. J. D. (2021). Deeply felt affect: The emergence of valence in deep active inference. *Neural Computation*, 33(2):398–446.
- Heydrich, L., Aspell, J. E., Marillier, G., Lavanchy, T., Herbelin, B., and Blanke, O. (2018). Cardio-visual full body illusion alters bodily self-consciousness and tactile processing in somatosensory cortex. *Scientific Reports*, 8(1):9230.
- Hickok, G., Houde, J., and Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, 69(3):407–422.
- Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5):393–402.
- Higgins, J. D. (1971). Set and uncertainty as factors influencing anticipatory cardiovascular responding in humans. *Journal of Comparative & Physiological Psychology*, 74(2):272–283.
- Hohwy, J. (2016). The self-evidencing brain. *Noûs*, 50(2):259–285.

- Hohwy, J. (2017). How to entrain your evil demon. In Metzinger, T. and Wiese, W., editors, *Philosophy and Predictive Processing*, chapter 2, pages 1–15. Frankfurt am Main: MIND Group.
- Hohwy, J. (2018). The predictive processing hypothesis. In Newen, A., de Bruin, L., and Gallagher, S., editors, *The Oxford handbook of 4E cognition*, chapter 7, pages 129–145. Oxford: Oxford University Press.
- Hohwy, J. (2020a). New directions in predictive processing. *Mind & Language*, 35(2):209–223.
- Hohwy, J. (2020b). Self-supervision, normativity and the free energy principle. *Synthese*.
- Hohwy, J. (2021). Conscious self-evidencing. *Review of Philosophy & Psychology*.
- Hohwy, J. and Michael, J. (2017). Why should any body have a self? In de Vignemont, F. and Alsmith, A. J. T., editors, *The subject’s matter: Self-consciousness and the body*, chapter 16, pages 363–391. Cambridge, MA: MIT Press.
- Holdgraf, C. R., de Heer, W., Pasley, B., Rieger, J., Crone, N., Lin, J. J., Knight, R. T., and Theunissen, F. E. (2016). Rapid tuning shifts in human auditory cortex enhance speech intelligibility. *Nature Communications*, 7:13654.
- Holroyd, C. B. and Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4):679–709.
- Hornberger, L. K. and Sahn, D. J. (2007). Rhythm abnormalities of the fetus. *Heart*, 93(10):1294–1300.
- Huang, Y. and Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5):580–593.
- Huetting, F. (2015). Four central questions about prediction in language processing. *Brain Research*, 1626:118–135.
- Huetting, F. and Mani, N. (2016). Is prediction necessary to understand language? probably not. *Language, Cognition & Neuroscience*, 31(1):19–31.
- Hurley, S. (1998). *Consciousness in action*. Cambridge, MA: Harvard University Press.
- Hurley, S. (2010). The varieties of externalism. In Menary, R., editor, *The extended mind*, chapter 6, pages 101–153. Cambridge, MA: MIT Press.
- Hutt, S., Krasich, K., Mills, C., Bosch, N., White, S., Brockmole, J. R., and D’Mello, S. K. (2019). Automated gaze-based mind wandering detection during computerized learning in classrooms. *User Modeling & User-Adapted Interaction*, 29:821–867.

- Hutto, D. D. and Myin, E. (2013). Neural representations not needed - no more pleas, please. *Phenomenology & the Cognitive Sciences*, 13(2):241–256.
- Irving, Z. C. (2016). Mind-wandering is unguided attention: Accounting for the “purposeful” wanderer. *Philosophical Studies*, 173:547–571.
- James, W. (1884). What is an emotion? *Mind*, 9:188–205.
- Jennings, J. R. (1986). Bodily changes during attention. In Coles, M. G. H., Donchin, E., and Porges, S. W., editors, *Psychophysiology: Systems, Processes, and Applications*, chapter 13, pages 268–289. New York & London: Guilford Press.
- Jennings, J. R. (1992). Is it important that the mind is in a body? inhibition and the heart. *Psychophysiology*, 29(4):369–383.
- Jennings, J. R., Averill, J. R., Opton, E. M., and Lazarus, R. S. (1970). Some parameters of heart rate change: Perceptual versus motor task requirements, noxiousness, and uncertainty. *Psychophysiology*, 7(2):194–212.
- Jennings, J. R. and van der Molen, M. W. (2002). Cardiac timing and the central regulation of action. *Psychological Research*, 66(4):337–349.
- Jennings, J. R. and van der Molen, M. W. (2005). Preparation for speeded action as a psychophysiological concept. *Psychological Bulletin*, 131(3):434–459.
- Jennings, J. R., van der Molen, M. W., Brock, K., and Somsen, R. J. (1992). On the synchrony of stopping motor responses and delaying heartbeats. *Journal of Experimental Psychology: Human Perception & Performance*, 18(2):422–436.
- Jennings, J. R., van der Molen, M. W., and Tanase, C. (2009). Preparing hearts and minds: Cardiac slowing and a cortical inhibitory network. *Psychophysiology*, 46(6):1170–1178.
- Jennings, J. R. and Wood, C. C. (1977). Cardiac cycle time effects on performance, phasic cardiac responses, and their intercorrelation in choice reaction time. *Psychophysiology*, 14(3):297–307.
- Jensen, O. and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, 4(186):1–8.
- Joffily, M. and Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLoS Computational Biology*, 9(6):e1003094.
- Johnson, K. A., Kelly, S. P., Bellgrove, M. A., Barry, E., Cox, M., Gill, M., and Robertson, I. H. (2007a). Response variability in attention deficit hyperactivity disorder: Evidence for neuropsychological heterogeneity. *Neuropsychologia*, 45(4):630–638.

- Johnson, K. A., Robertson, I. H., Kelly, S. P., Silk, T. J., Barry, E., Dáibhis, A., Watchorn, A., Keavey, M., Fitzgerald, M., Gallagher, L., Gill, M., and Bellgrove, M. A. (2007b). Dissociation in performance of children with adhd and high-functioning autism on a task of sustained attention. *Neuropsychologia*, 45(10):2234–2245.
- Kagan, J. and Lewis, M. (1965). Studies of attention in the human infant. *Merrill-Palmer Quarterly of Behavior & Development*, 11(2):95–127.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, N.J.: Prentice-Hall, Inc.
- Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language & Linguistics Compass*, 2(4):647–670.
- Kandler, K., Clause, A., and Noh, J. (2009). Tonotopic reorganization of developing auditory brainstem circuits. *Nature Neuroscience*, 12(6):711–717.
- Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., and Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological Science*, 18(7):614–621.
- Kastner, L., Kube, J., Villringer, A., and Neumann, J. (2017). Cardiac concomitants of feedback and prediction error processing in reinforcement learning. *Frontiers in Neuroscience*, 11:598.
- Katz, L. C. and Shatz, C. J. (1996). Synaptic activity and the construction of cortical circuits. *Science*, 274(5290):1133–1138.
- Kayser, S. J., Ince, R. A. A., Gross, J., and Kayser, C. (2015). Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. *Journal of Neuroscience*, 35(44):14691–14701.
- Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *Journal of Neuroscience*, 30(2):620–628.
- Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, J. S., Feinstein, J. S., Feusner, J. D., Garfinkel, S. N., Lane, R. D., Mehling, W. E., Meuret, A. E., Nemeroff, C. B., Oppenheimer, S., Petzschner, F. H., Pollatos, O., Rhudy, J. L., Schramm, L. P., Simmons, W. K., Stein, M. B., Stephan, K. E., Van Den Bergh, O., Van Diest, I., von Leupoldt, A., and Paulus, M. P. (2018). Interoception and mental health: A roadmap. *Biological Psychiatry: Cognitive Neuroscience & Neuroimaging*, 3:501–513.

- Killingsworth, M. A. and Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, 330(6006):932.
- Kingma, E. (2019). Were you a part of your mother? *Mind*, 128:609–646.
- Kirchhoff, M. D. (2015). Extended cognition and the causal-constitutive fallacy: In search for a diachronic and dynamical conception of constitution. *Philosophy & Phenomenological Research*, 90(2):320–360.
- Kirchhoff, M. D. and Kiverstein, J. (2019). How to determine the boundaries of the mind: A markov blanket proposal. *Synthese*.
- Kirkby, L. A., Sack, G. S., Firl, A., and Feller, M. B. (2013). A role for correlated spontaneous activity in the assembly of neural circuits. *Neuron*, 80(5):1129–1144.
- Kiverstein, J. and Miller, M. (2015). The embodied brain: Towards a radical embodied cognitive neuroscience. *Frontiers in Human Neuroscience*, 9(237):1–11.
- Kiverstein, J. and Sims, M. (2021). Is free-energy minimisation the mark of the cognitive? *Biology & Philosophy*, 36(25).
- Kleinschmidt, D. F. and Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2):148–203.
- Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, 16(12):606–617.
- Klimesch, W. (2018). The frequency architecture of brain and brain body oscillations: an analysis. *European Journal of Neuroscience*, 48(7):2431–2453.
- Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). Eeg alpha oscillations: The inhibition-timing hypothesis. *Brain Research Reviews*, 53(1):63–88.
- Knill, D. C. and Pouget, A. (2004). The bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Cognitive Sciences*, 27(12):712–719.
- Komatsu, H. (2006). The neural mechanisms of perceptual filling-in. *Nature Reviews Neuroscience*, 7(3):220–231.
- Kösem, A. and van Wassenhove, V. (2017). Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. *Language, Cognition & Neuroscience*, 32(5):536–544.
- Köster, M., Kayhan, E., Langeloh, M., and Hoehl, S. (2020). Making sense of the world: Infant learning from a predictive processing perspective. *Perspectives on Psychological Science*, 15(3):562–571.

- Kunzendorf, S., Klotzsche, F., Akbal, M., Villringer, A., Ohl, S., and Gaebler, M. (2019). Active information sampling varies across the cardiac cycle. *Psychophysiology*, 56(5):e13322.
- Kuperberg, G. R. (2016). Separate streams or probabilistic inference? what the n400 can tell us about the comprehension of events. *Language, Cognition & Neuroscience*, 31(5):602–616.
- Kuperberg, G. R. and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition & Neuroscience*, 31(1):32–59.
- Kutas, M., DeLong, K. A., and Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In Bar, M., editor, *Predictions in the brain: Using our past to generate a future*, chapter 15, pages 190–207. OUP.
- Kutas, M. and Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the n400 component of the event-related brain potential (erp). *Annual Review of Psychology*, 62:621–647.
- Kvasov, D. G. and Korovina, M. V. (1965). The reflex organization of perception and the proprio-muscular apparatus of the analyzers (of the sense organs). In Voronin, L. G., Leontiev, A. R., Luria, A. R., Sokolov, E. N., and Vinogradova, O. S., editors, *Orienting reflex and exploratory behavior*, pages 178–186. Washington, D.C.: American Institute of Biological Sciences.
- Lacey, B. C. and Lacey, J. I. (1974). Studies of heart rate and other bodily processes in sensorimotor behavior. In Obrist, P. A., Black, A. H., Brener, J., and DiCara, L. V., editors, *Cardiovascular psychophysiology: Current issues in response mechanisms, biofeedback and methodology*, chapter 26, pages 538–564. Chicago, IL: Aldine Publishing Co.
- Lacey, B. C. and Lacey, J. I. (1977). Change in heart period: A function of sensorimotor event timing within the cardiac cycle. *Physiological Psychology*, 5(3):383–393.
- Lacey, B. C. and Lacey, J. I. (1978). Two-way communication between the heart and the brain: Significance of time within the cardiac cycle. *American Psychologist*, 33(2):99–113.
- Lacey, B. C. and Lacey, J. I. (1980). Presidential address, 1979. cognitive modulation of time-dependent primary bradycardia. *Psychophysiology*, 17(3):209–221.
- Lacey, J. I. (1959). Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In Rubinstein, E. A. and Parloff, M. B., editors, *Research*

- in *Psychotherapy*, pages 160–208. Washington, D.C.: American Psychological Association.
- Lacey, J. I. (1967). Somatic response patterning and stress: Some revisions of activation theory. In Appley, M. H. and Trumbull, R., editors, *Psychological Stress: Issues in Research*, chapter 2, pages 14–37. New York, NY: Appleton-Century-Crofts.
- Lacey, J. I. (1972). Some cardiovascular correlates of sensorimotor behavior: Examples of visceral afferent feedback? In Hockman, C. H., editor, *Limbic system mechanisms and autonomic function*, chapter 11, pages 175–196. Springfield, IL: Charles C. Thomas.
- Lacey, J. I., Kagan, J., Lacey, B. C., and Moss, H. A. (1963). The visceral level: Situational determinants and behavioral correlates of autonomic response patterns. In Knapp, P. H., editor, *Expression of the emotions in man*, chapter 9, pages 161–196. New York, NY: International Universities Press.
- Lacey, J. I. and Lacey, B. C. (1958). The relationship of resting autonomic activity to motor impulsivity. In *Proceedings of the Association for Research in Nervous & Mental Disease*, volume 36, pages 144–209.
- Lacey, J. I. and Lacey, B. C. (1970). Some autonomic-central nervous system interrelationships. In Black, P., editor, *Physiological correlates of emotion*, chapter 10, pages 205–227. New York & London: Academic Press.
- Lakatos, P., Barczak, A., Neymotin, S. A., McGinnis, T., Ross, D., Javitt, D. C., and O’Connell, M. N. (2016). Global dynamics of selective attention and its lapses in primary auditory cortex. *Nature Neuroscience*, 19(12):1707–1717.
- Lakatos, P., Gross, J., and Thut, G. (2019). A new unifying account of the roles of neuronal entrainment. *Current Biology*, 29(18):R890–R905.
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, 320(5872):110–113.
- Lakatos, P., Musacchia, G., O’Connell, M. N., Falchier, A. Y., Javitt, D. C., and Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron*, 77(4):750–761.
- Lakatos, P., O’Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., and Schroeder, C. E. (2009). The leading sense: Supramodal control of neurophysiological context by attention. *Neuron*, 64(3):419–430.
- Lakoff, G. and Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

- Lakoff, G. and Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York, NY: Basic Books.
- Lam, N. H. L., Schoffelen, J.-M., Uddén, J., Hultén, A., and Hagoort, P. (2016). Neural activity during sentence processing as reflected in theta, alpha, beta, and gamma oscillations. *NeuroImage*, 142:43–54.
- Laming, D. (1979). Choice reaction performance following an error. *Acta Psychologica*, 43:199–224.
- Lancaster, M. A., Renner, M., Martin, C.-A., Wenzel, D., Bicknell, L. S., Hurles, M. E., Homfray, T., Penninger, J. M., Jackson, A. P., and Knoblich, J. A. (2013). Cerebral organoids model human brain development and microcephaly. *Nature*, 501(7467):373–379.
- Lang, P. J., Öhman, A., and Simons, R. F. (1978). The psychophysiology of anticipation. In Requin, J., editor, *Attention and performance VII*, chapter 25, pages 469–485. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Lange, C. G. (1922). The emotions: A psychophysiological study. In Dunlap, K., editor, *The emotions*, pages 33–90. Baltimore, M.D.: Williams & Wilkins Co.
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition & Emotion*, 24(8):1377–1388.
- Lappin, J. S. and Eriksen, C. W. (1966). Use of a delayed signal to stop a visual reaction-time response. *Journal of Experimental Psychology*, 72(6):805–811.
- Large, E. W. and Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106(1):119–159.
- Larra, M. F., Finke, J. B., Wascher, E., and Schächinger, H. (2020). Disentangling sensorimotor and cognitive cardioafferent effects: A cardiac-cycle-time study on spatial stimulus-response compatibility. *Scientific Reports*, 10(1):4059.
- Leaver, A. M. and Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *Journal of Neuroscience*, 30(22):7604–7612.
- Lee, H. and Noppeney, U. (2011). Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. *Journal of Neuroscience*, 31(31):11338–11350.

- Legendre, G., Andrillon, T., Koroma, M., and Kouider, S. (2019). Sleepers track informative speech in a multitalker environment. *Nature Human Behaviour*, 3(3):274–283.
- Leighton, A. H. and Lohmann, C. (2016). The wiring of developing sensory circuits – from patterned spontaneous activity to synaptic plasticity mechanisms. *Frontiers in Neural Circuits*, 10:71.
- Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: Manipulating bodily self-consciousness. *Science*, 317(5841):1096–1099.
- Lenth, R. (2020). emmeans: Estimated marginal means, aka least-squares means.
- Leonard, M. K., Baud, M. O., Sjerps, M. J., and Chang, E. F. (2016). Perceptual restoration of masked speech in human cortex. *Nature Communications*, 7:13619.
- Levelt, W. J. M. (1965). *On binocular rivalry*. Soesterberg: Institute for Perception.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.
- Lewis, A. G. and Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex*, 68:155–168.
- Liebherr, M., Corcoran, A. W., Alday, P. M., Coussens, S., Bellan, V., Howlett, C. A., Immink, M. A., Kohler, M. J., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2021). Eeg and behavioral correlates of attentional processing while walking and navigating naturalistic environments. *bioRxiv*.
- Limanowski, J. and Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Frontiers in Human Neuroscience*, 7(547):1–12.
- Limanowski, J. and Friston, K. J. (2018). ‘seeing the dark’: Grounding phenomenal transparency and opacity in precision estimation for active inference. *Frontiers in Psychology*, 9:643.
- Linson, A., Parr, T., and Friston, K. J. (2020). Active inference, stressors, and psychological trauma: A neuroethological model of (mal)adaptive explore-exploit dynamics in ecological context. *Behavioural Brain Research*, 380:112421.
- Loewy, A. D. (1981). Descending pathways to sympathetic and parasympathetic preganglionic neurons. *Journal of the Autonomic Nervous System*, 3(2–4):265–275.
- Logan, G. D. and Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, 91(3):295–327.

- Lüdtke, D. (2018). *ggeffects*: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, 3(26):772.
- Lüdtke, D., Ben-Shachar, M., Patil, I., Waggoner, P., and Makowski, D. (2021). *performance*: An r package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60):3139.
- Luhmann, H. J., Sinning, A., Yang, J.-W., Reyes-Puerta, V., Stüttgen, M. C., Kirischuk, S., and Kilb, W. (2016). Spontaneous neuronal activity in developing neocortical networks: From single cells to large-scale interactions. *Frontiers in Neural Circuits*, 10:40.
- Lukowska, M., Sznajder, M., and Wierzchoń, M. (2018). Error-related cardiac response as information for visibility judgements. *Scientific Reports*, 8(1):1131.
- Luo, H. and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54(6):1001–1010.
- Macefield, V. G. (2003). Cardiovascular and respiratory modulation of tactile afferents in the human finger pad. *Experimental Physiology*, 88(5):617–625.
- Mackworth, N. H. (1956). Vigilance. *Nature*, 178(4547):1375–1377.
- Macmillan, N. A. and Creelman, C. D. (2005). *Detection theory: A user’s guide*. New York, NY: Psychological Press, 2nd edition.
- Manly, T., Anderson, V., Nimmo-Smith, I., Turner, A., Watson, P., and Robertson, I. H. (2001). The differential assessment of children’s attention: The test of everyday attention for children (tea-ch), normative sample and adhd performance. *Journal of Child Psychology & Psychiatry*, 42(8):1065–1081.
- Manly, T., Davison, B., Heutink, J., Galloway, M., and Robertson, I. H. (2000). Not enough time or not enough attention? speed, error and self-maintained control in the sustained attention to response test (sart). *Clinical Neuropsychological Assessment*, 3:167–177.
- Manly, T., Owen, A. M., McAvinue, L., Datta, A., Lewis, G. H., Scott, S. K., Rorden, C., Pickard, J., and Robertson, I. H. (2003). Enhancing the sensitivity of a sustained attention task to frontal damage: Convergent clinical and functional imaging evidence. *Neurocase*, 9(4):340–349.
- Manly, T., Robertson, I. H., Galloway, M., and Hawkins, K. (1999). The absent mind: Further investigations of sustained attention to response. *Neuropsychologia*, 37(6):661–670.

- Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology*, 49(4):549–565.
- Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of eeg- and meg-data. *Journal of Neuroscience Methods*, 164(1):177–190.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman.
- Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, 189(4198):226–228.
- Marslen-Wilson, W. D. and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10:29–63.
- Martínez Quintero, A. and De Jaegher, H. (2020). Pregnant agencies: Movement and participation in maternal-fetal interactions. *Frontiers in Psychology*, 11:1977.
- Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language & Cognitive Processes*, 27(7-8):953–978.
- McClelland, J. L., Mirman, D., and Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, 10(8):363–369.
- McIntyre, D., Ring, C., Edwards, L., and Carroll, D. (2008). Simple reaction time as a function of the phase of the cardiac cycle in young adults at risk for hypertension. *Psychophysiology*, 45(2):333–336.
- McIntyre, D., Ring, C., Hamer, M., and Carroll, D. (2007). Effects of arterial and cardiopulmonary baroreceptor activation on simple and choice reaction times. *Psychophysiology*, 44(6):874–879.
- McMahon, C. M., Boisvert, I., de Lissa, P., Granger, L., Ibrahim, R., Lo, C. Y., Miles, K., and Graham, P. L. (2016). Monitoring alpha oscillations and pupil dilation across a performance-intensity function. *Frontiers in Psychology*, 7:745.
- McVay, J. C. and Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 35(1):196–204.
- Menary, R., editor (2010a). *The extended mind*. Cambridge, MA: MIT Press.
- Menary, R. (2010b). The holy grail of cognitivism: A response to adams and aizawa. *Phenomenology & the Cognitive Sciences*, 9:605–618.

- Mesgarani, N. and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485(7397):233–236.
- Metzinger, T. (2003). *Being no one: The self-model theory of subjectivity*. Cambridge, MA: MIT Press.
- Metzinger, T. (2013). The myth of cognitive agency: Subpersonal thinking as a cyclically recurring loss of mental autonomy. *Frontiers in Psychology*, 4:931.
- Metzinger, T. (2017). The problem of mental action: Predictive control without sensory sheets. In Metzinger, T. and Wiese, W., editors, *Philosophy and Predictive Processing*, chapter 19, pages 1–26. Frankfurt am Main: MIND Group.
- Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, 48(7):2609–2621.
- Mildner, J. N. and Tamir, D. I. (2019). Spontaneous thought as an unconstrained memory process. *Trends in Neurosciences*, 42(11):763–777.
- Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., de Lissa, P., Graham, P., and Lyxell, B. (2017). Objective assessment of listening effort: Coregistration of pupillometry and eeg. *Trends in Hearing*, 21:1–13.
- Milkowski, M. (2019). Embodied cognition. In Sprevak, M. and Colombo, M., editors, *The routledge handbook of the computational mind*, chapter 24, pages 323–338. Oxon & New York: Routledge.
- Miller, G. A. and Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning & Verbal Behavior*, 2:217–228.
- Mirza, M. B., Adams, R. A., Mathys, C. D., and Friston, K. J. (2016). Scene construction, visual foraging, and active inference. *Frontiers in Computational Neuroscience*, 10(56):1–16.
- Mittner, M., Hawkins, G. E., Boekel, W., and Forstmann, B. U. (2016). A neural model of mind wandering. *Trends in Cognitive Sciences*, 20(8):570–578.
- Mizumoto, M. and Ishikawa, M. (2005). Immunity to error through misidentification and the bodily illusion experiment. *Journal of Consciousness Studies*, 12(7):3–19.
- Monti, A., Porciello, G., Tieri, G., and Aglioti, S. M. (2020). The ”embreathment” illusion highlights the role of breathing in corporeal awareness. *Journal of Neurophysiology*, 123(1):420–427.

- Montirossi, R. and McGlone, F. (2020). The body comes first: Embodied reparation and the co-creation of infant bodily-self. *Neuroscience & Biobehavioral Reviews*, 113:77–87.
- Mooney, R., Penn, A. A., Gallego, R., and Shatz, C. J. (1996). Thalamic relay of spontaneous retinal activity prior to vision. *Neuron*, 17(5):863–874.
- Morillon, B., Hackett, T. A., Kajikawa, Y., and Schroeder, C. E. (2015). Predictive motor control of sensory dynamics in auditory active sensing. *Current Opinion in Neurobiology*, 31:230–238.
- Morillon, B., Schroeder, C. E., and Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nature Communications*, 5(5255):1–9.
- Müller, F. and O’Rahilly, R. (2006). The amygdaloid complex and the medial and lateral ventricular eminences in staged human embryos. *Journal of Anatomy*, 208(5):547–564.
- Nakamura, N. H., Fukunaga, M., and Oku, Y. (2018). Respiratory modulation of cognitive performance during the retrieval process. *PLoS One*, 13(9):e0204021.
- Nave, K., Deane, G., Miller, M., and Clark, A. (2020). Wilding the predictive brain. *Wiley Interdisciplinary Reviews: Cognitive Science*, 11(6):e1542.
- Niemi, P. and Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin*, 89(1):133–162.
- Nieuwland, M. S. (2019). Do ‘early’ brain responses reveal word form prediction during language comprehension? a critical review. *Neuroscience & Biobehavioral Reviews*, 96:367–400.
- Nobre, A. C., Correa, A., and Coull, J. T. (2007). The hazards of time. *Current Opinion in Neurobiology*, 17(4):465–470.
- Nobre, A. C. and van Ede, F. (2018). Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience*, 19(1):34–48.
- Noë, A. (2004). *Action in perception*. Cambridge, MA: MIT Press.
- Noë, A. (2009). *Out of our heads: Why you are not your brain, and other lessons from the biology of consciousness*. New York, NY: Hill & Wang.
- Norman, D. A. and Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In Davidson, R. J., Schwartz, G. E., and Shapiro, D. E., editors, *Consciousness and self-regulation*, pages 1–14. New York, NY: Plenum Press.
- Notebaert, W., Houtman, F., Opstal, F. V., Gevers, W., Fias, W., and Verguts, T. (2009). Post-error slowing: An orienting account. *Cognition*, 111(2):275–279.

- Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., Howard, 3rd, M. A., and Brugge, J. F. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *Journal of Neuroscience*, 29(49):15564–15574.
- Nowlin, J. B., Eisdorfer, C., Whalen, R., and Troyer, W. G. (1970). The effect of exogenous changes in heart rate and rhythm upon reaction time performance. *Psychophysiology*, 7(2):186–193.
- Obleser, J. and Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences*, 23(11):913–926.
- Obleser, J. and Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex*, 20(3):633–640.
- Obleser, J. and Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cerebral Cortex*, 22(11):2466–2477.
- Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., and Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *Journal of Neuroscience*, 32(36):12376–12383.
- Obrist, P. A. (1963). Cardiovascular differentiation of sensory stimuli. *Psychosomatic Medicine*, 25:450–459.
- Obrist, P. A. (1968). Heart rate and somatic-motor coupling during classical aversive conditioning in humans. *Journal of Experimental Psychology*, 77(2):180–193.
- Obrist, P. A. (1976). The cardiovascular-behavioral interaction – as it appears today. *Psychophysiology*, 13(2):95–107.
- Obrist, P. A. (1981). *Cardiovascular psychophysiology: A perspective*. New York & London: Plenum Press.
- Obrist, P. A. and Webb, R. A. (1967). Heart rate during conditioning in dogs: Relationship to somatic-motor activity. *Psychophysiology*, 4(1):7–34.
- Obrist, P. A., Webb, R. A., and Sutterer, J. R. (1969). Heart rate and somatic changes during aversive conditioning and a simple reaction time task. *Psychophysiology*, 5(6):696–723.
- Obrist, P. A., Webb, R. A., Sutterer, J. R., and Howard, J. L. (1970). The cardiac-somatic relationship: Some reformulations. *Psychophysiology*, 6(5):569–587.
- Ohl, S., Wohltat, C., Kliegl, R., Pollatos, O., and Engbert, R. (2016). Microsaccades are coupled to heartbeat. *Journal of Neuroscience*, 36(4):1237–1241.

- Öhman, A., Hamm, A., and Hugdahl, K. (2000). Cognition and the autonomic nervous system: Orienting, anticipation, and conditioning. In Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G., editors, *Handbook of psychophysiology*, chapter 20, pages 533–575. Cambridge: Cambridge University Press, 2nd edition.
- O’Keeffe, F. M., Dockree, P. M., and Robertson, I. H. (2004). Poor insight in traumatic brain injury mediated by impaired error processing? evidence from electrodermal activity. *Brain Research: Cognitive Brain Research*, 22(1):101–112.
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). Fieldtrip: Open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational Intelligence & Neuroscience*, 2011:156869.
- O’Regan, J. K. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral & Brain Sciences*, 24(5):939–1031.
- O’Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., and Lalor, E. C. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial eeg. *Cerebral Cortex*, 25(7):1697–1706.
- Ottaviani, C., Medea, B., Lonigro, A., Tarvainen, M., and Couyoumdjian, A. (2015a). Cognitive rigidity is mirrored by autonomic inflexibility in daily life perseverative cognition. *Biological Psychology*, 107:24–30.
- Ottaviani, C., Shahabi, L., Tarvainen, M., Cook, I., Abrams, M., and Shapiro, D. (2015b). Cognitive, behavioral, and autonomic correlates of mind wandering and perseverative cognition in major depression. *Frontiers in Neuroscience*, 8:433.
- Ottaviani, C., Shapiro, D., and Couyoumdjian, A. (2013). Flexibility as the key for somatic health: From mind wandering to perseverative cognition. *Biological Psychology*, 94(1):38–43.
- Owens, A. P., Allen, M., Ondobaka, S., and Friston, K. J. (2018). Interoceptive inference: From computational neuroscience to clinic. *Neuroscience & Biobehavioral Reviews*, 90:174–183.
- Panksepp, J. and Northoff, G. (2009). The trans-species core self: The emergence of active cultural and neuro-ecological agents through self-related processing within subcortical-cortical midline networks. *Consciousness & Cognition*, 18:193–215.
- Parasuraman, R. (1979). Memory load and event rate control sensitivity decrements in sustained attention. *Science*, 205(4409):924–927.

- Park, H.-D., Barnoud, C., Trang, H., Kannape, O. A., Schaller, K., and Blanke, O. (2020). Breathing is coupled with voluntary action and the cortical readiness potential. *Nature Communications*, 11(1):289.
- Park, H.-D., Bernasconi, F., Salomon, R., Tallon-Baudry, C., Spinelli, L., Seeck, M., Schaller, K., and Blanke, O. (2018). Neural sources and underlying mechanisms of neural responses to heartbeats, and their role in bodily self-consciousness: An intracranial eeg study. *Cerebral Cortex*, 28(7):2351–2364.
- Park, H.-D. and Tallon-Baudry, C. (2014). The neural subjective frame: From bodily signals to perceptual consciousness. *Philosophical Transactions of the Royal Society B*, 369(20130208):1–9.
- Parr, T., Corcoran, A. W., Friston, K. J., and Hohwy, J. (2019). Perceptual awareness and active inference. *Neuroscience of Consciousness*, 5(1):niz012.
- Parr, T. and Friston, K. J. (2017). The active construction of the visual world. *Neuropsychologia*, 104:92–101.
- Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T., and Chang, E. F. (2012). Reconstructing speech from human auditory cortex. *PLoS Biology*, 10(1):e1001251.
- Pattyn, N., Neyt, X., Henderickx, D., and Soetens, E. (2008). Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue? *Physiology & Behavior*, 93(1-2):369–378.
- Patzelt, E. H., Hartley, C. A., and Gershman, S. J. (2018). Computational phenotyping: Using models to understand individual differences in personality, development, and mental illness. *Personality Neuroscience*, 1:e18.
- Paulus, M. P., Feinstein, J. S., and Khalsa, S. S. (2019). An active inference approach to interoceptive psychopathology. *Annual Review of Clinical Psychology*, 15:97–122.
- Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Oxford: Oxford University Press.
- Peebles, D. and Bothell, D. (2004). Modelling performance in the sustained attention to response task. In *Proceedings of the Sixth International Conference on Cognitive Modeling*, pages 231–236. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Peelle, J. E. and Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, 3:320.

- Peelle, J. E., Gross, J., and Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb Cortex*, 23(6):1378–1387.
- Penn, A. A. and Shatz, C. J. (1999). Brain waves and brain wiring: The role of endogenous and sensory-driven neural activity in development. *Pediatric Research*, 45(4 Pt 1):447–458.
- Perl, O., Ravia, A., Rubinson, M., Eisen, A., Soroka, T., Mor, N., Secundo, L., and Sobel, N. (2019). Human non-olfactory cognition phase-locked with inhalation. *Nature Human Behaviour*, 3(5):501–512.
- Perrykkad, K. and Hohwy, J. (2020). Fidgeting as self-evidencing: A predictive processing account of non-goal-directed action. *New Ideas in Psychology*, 56(100750):1–8.
- Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, 9(2):148–158.
- Pessoa, L. and De Weerd, P., editors (2003). *Filling-in: From perceptual completion to cortical reorganization*. Oxford: Oxford University Press.
- Peters, A., McEwen, B. S., and Friston, K. J. (2017). Uncertainty and stress: Why it causes diseases and how it is mastered by the brain. *Progress in Neurobiology*, 156:164–188.
- Petkov, C. I. and Sutter, M. L. (2011). Evolutionary conservation and neuronal mechanisms of auditory perceptual restoration. *Hearing Research*, 271(1-2):54–65.
- Petzschnner, F. H., Garfinkel, S. N., Paulus, M. P., Koch, C., and Khalsa, S. S. (2021). Computational models of interoception and body regulation. *Trends in Neurosciences*, 44(1):63–76.
- Petzschnner, F. H., Weber, L. A., Wellstein, K. V., Paolini, G., Do, C. T., and Stephan, K. E. (2019). Focus of attention modulates the heartbeat evoked potential. *NeuroImage*, 186:595–606.
- Pezzulo, G. (2012). An active inference view of cognitive control. *Frontiers in Psychology*, 3(478):1–2.
- Pezzulo, G. (2014). Why do you fear the bogeyman? an embodied predictive coding model of perceptual inference. *Cognitive, Affective, & Behavioral Neuroscience*, 14(3):902–911.
- Pezzulo, G. (2017). Tracing the roots of cognition in predictive processing. In Metzinger, T. and Wiese, W., editors, *Philosophy and Predictive Processing*, chapter 20, pages 1–20. Frankfurt am Main: MIND Group.

- Pezzulo, G. and Castelfranchi, C. (2009). Thinking as the control of imagination: A conceptual framework for goal-directed systems. *Psychological Research*, 73(4):559–577.
- Pezzulo, G. and Cisek, P. (2016). Navigating the affordance landscape: Feedback control as a process model of behavior and cognition. *Trends in Cognitive Sciences*, 20(6):414–424.
- Pezzulo, G., Iodice, P., Barca, L., Chausse, P., Monceau, S., and Mermillod, M. (2018). Increased heart rate after exercise facilitates the processing of fearful but not disgusted faces. *Scientific Reports*, 8(1):398.
- Pezzulo, G., Rigoli, F., and Friston, K. J. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, 134:17–35.
- Pham, P. and Wang, J. (2015). Attentivelearner: Improving mobile mooc learning via implicit heart rate monitoring. In Conati, C., Heffernan, N., Mitrovic, A., and Verdejo, M., editors, *Artificial intelligence in education. AIED2015.*, volume 9112 of *Lecture Notes in Computer Science*, pages 367–376. Switzerland: Springer International Publishing.
- Piai, V., Anderson, K. L., Lin, J. J., Dewar, C., Parvizi, J., Dronkers, N. F., and Knight, R. T. (2016). Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. *Proceedings of the National Academy of Sciences*, 113(40):11366–11371.
- Pickering, M. J. and Clark, A. (2014). Getting ahead: Forward models and their place in cognitive architecture. *Trends in Cognitive Sciences*, 18(9):451–456.
- Pickering, M. J. and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral & Brain Sciences*, 36(4):329–392.
- Pillai, M. and James, D. (1990). The development of fetal heart rate patterns during normal pregnancy. *Obstetrics & Gynecology*, 76(5 Pt 1):812–816.
- Poeppel, D. and Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, 21(6):322–334.
- Poeppel, D., Idsardi, W. J., and van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493):1071–1086.
- Poeppel, D. and Monahan, P. J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language & Cognitive Processes*, 26(7):935–951.

- Porges, S. W. (1972). Heart rate variability and deceleration as indexes of reaction time. *Journal of Experimental Psychology*, 92(1):103–110.
- Porges, S. W. (1992). Autonomic regulation and attention. In Campbell, B. A., Hayne, H., and Richardson, R., editors, *Attention and information processing in infants and adults: Perspectives from human and animal research*, chapter 8, pages 201–223. New York and London: Psychology Press.
- Porges, S. W. and Raskin, D. C. (1969). Respiratory and heart rate components of attention. *Journal of Experimental Psychology*, 81(3):497–503.
- Posner, M. I., Rueda, R., and Kanske, P. (2007). Probing the mechanisms of attention. In Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G., editors, *Handbook of psychophysiology*, chapter 18, pages 410–432. Cambridge: Cambridge University Press, 3rd edition.
- Pramme, L., Larra, M. F., Schächinger, H., and Frings, C. (2014). Cardiac cycle time effects on mask inhibition. *Biological Psychology*, 100:115–121.
- Pramme, L., Larra, M. F., Schächinger, H., and Frings, C. (2016). Cardiac cycle time effects on selection efficiency in vision. *Psychophysiology*, 53(11):1702–1711.
- Pribram, K. H. and McGuinness, D. (1975). Arousal, activation, and effort in the control of attention. *Psychological Review*, 82(2):116–149.
- Prinz, J. (2009). Is consciousness embodied? In Robbins, P. and Aydede, M., editors, *The Cambridge handbook of situated cognition*, chapter 22, pages 419–436. Cambridge: Cambridge University Press.
- Putnam, H. (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.
- Qian, X., Song, H., and Ming, G.-L. (2019). Brain organoids: Advances, applications and challenges. *Development*, 146(8):dev166074.
- Quadt, L., Critchley, H. D., and Garfinkel, S. N. (2018). The neurobiology of interoception in health and disease. *Annals of the New York Academy of Sciences*, 1428(1):112–128.
- Quattrocki, E. and Friston, K. J. (2014). Autism, oxytocin and interoception. *Neuroscience & Biobehavioral Reviews*, 47:410–430.

- Quelhas Martins, A., McIntyre, D., and Ring, C. (2014). Effects of baroreceptor stimulation on performance of the sternberg short-term memory task: A cardiac cycle time study. *Biological Psychology*, 103:262–266.
- Quigley, K. S., Kanoski, S., Grill, W. M., Barrett, L. F., and Tsakiris, M. (2021). Functions of interoception: From energy regulation to experience of the self. *Trends in Neurosciences*, 44(1):29–38.
- R Core Team (2019). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Rabbitt, P. M. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, 71(2):264–272.
- Rabbitt, P. M. (1968). Three kinds of error-signalling responses in a serial choice task. *Quarterly Journal of Experimental Psychology*, 20(2):179–188.
- Rae, C. L., Botan, V. E., Gould van Praag, C. D., Herman, A. M., Nyssönen, J. A. K., Watson, D. R., Duka, T., Garfinkel, S. N., and Critchley, H. D. (2018). Response inhibition on the stop signal task improves during cardiac contraction. *Scientific Reports*, 8(1):9136.
- Ramachandran, V. S. and Gregory, R. L. (1991). Perceptual filling in of artificially induced scotomas in human vision. *Nature*, 350(6320):699–702.
- Ramstead, M. J., Kirchhoff, M. D., and Friston, K. J. (2020). A tale of two densities: active inference is enactive inference. *Adapt Behav*, 28(4):225–239.
- Ramstead, M. J. D., Kirchhoff, M. D., Constant, A., and Friston, K. J. (2019). Multiscale integration: Beyond internalism and externalism. *Synthese*.
- Rao, R. P. N. and Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87.
- Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, 8(4):516–521.
- Rauschecker, J. P. and Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6):718–724.
- Reason, J. (1990). *Human error*. Cambridge: Cambridge University Press.
- Reichle, E. D., Reineberg, A. E., and Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science*, 21(9):1300–1310.

- Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science*, 212(4497):947–949.
- Ridderinkhof, K. R., van den Wildenberg, W. P. M., Segalowitz, S. J., and Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain & Cognition*, 56(2):129–140.
- Riecke, L., Esposito, F., Bonte, M., and Formisano, E. (2009). Hearing illusory sounds in noise: the timing of sensory-perceptual transformations in auditory cortex. *Neuron*, 64(4):550–561.
- Riecke, L., Vanbussel, M., Hausfeld, L., Başkent, D., Formisano, E., and Esposito, F. (2012). Hearing an illusory vowel in noise: suppression of auditory cortical activity. *Journal of Neuroscience*, 32(23):8024–8034.
- Rimmele, J. M., Zion Golumbic, E., Schröger, E., and Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex*, 68:144–154.
- Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., and Yiend, J. (1997). 'oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, 35(6):747–758.
- Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 336(1278):367–373.
- Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, Jr., E. D., and Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology*, 20(5):343–350.
- RStudio Team (2015). *RStudio: Integrated development for R*. RStudio, Inc., Boston, MA.
- Rumelhart, D. E. (1977). Toward an interactive model of reading. In Singer, H. and Ruddell, R. B., editors, *Theoretical models and processes of reading*, pages 722–750. Newark, DE: International Reading Association.
- Saari, M. J. and Pappas, B. A. (1976). Cardiac cycle phase and movement and reaction times. *Perceptual & Motor Skills*, 42(3):767–770.
- Saffran, J. R. and Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, 69:181–203.

- Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, 110(4):474–494.
- Sandman, C. A., McCanne, T. R., Kaiser, D. N., and Diamond, B. (1977). Heart rate and cardiac phase influences on visual perception. *Journal of Comparative & Physiological Psychology*, 91(1):189–202.
- Saper, C. B. (2002). The central autonomic nervous system: Conscious visceral perception and autonomic pattern generation. *Annual Review of Neuroscience*, 25:433–469.
- Sassenhagen, J. and Draschkow, D. (2019). Cluster-based permutation tests of meg/eeg data do not establish significance of effect latency or location. *Psychophysiology*, 56(6):e13335.
- Schachter, S. and Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69:379–399.
- Schroeder, C. E. and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1):9–18.
- Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., and Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology*, 20(2):172–176.
- Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H. B., Kronbichler, M., and Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *eLife*, 8.
- Scott, S. K. (2019). From speech and talkers to the social world: The neural processing of human spoken language. *Science*, 366(6461):58–62.
- Scott, S. K., McGettigan, C., and Eisner, F. (2009). A little more conversation, a little less action – candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, 10(4):295–302.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral & Brain Sciences*, 3:417–457.
- Segar, J. L. (1997). Ontogeny of the arterial and cardiopulmonary baroreflex during fetal and postnatal life. *American Journal of Physiology*, 273(2 Pt 2):R457–R471.
- Sel, A., Azevedo, R. T., and Tsakiris, M. (2017). Heartfelt self: Cardio-visual integration affects self-face recognition and interoceptive cortical processing. *Cerebral Cortex*, 27(11):5144–5155.

- Seli, P. (2016). The attention-lapse and motor decoupling accounts of sart performance are not mutually exclusive. *Consciousness & Cognition*, 41:189–198.
- Seli, P., Cheyne, J. A., and Smilek, D. (2012). Attention failures versus misplaced diligence: Separating attention lapses from speed-accuracy trade-offs. *Consciousness & Cognition*, 21(1):277–291.
- Seli, P., Kane, M. J., Smallwood, J., Schacter, D. L., Maillet, D., Schooler, J. W., and Smilek, D. (2018). Mind-wandering as a natural kind: A family-resemblances view. *Trends in Cognitive Sciences*, 22(6):479–490.
- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11):565–573.
- 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.
- Seth, A. K. (2015). The cybernetic bayesian brain: From interoceptive inference to sensorimotor contingencies. In Metzinger, T. and Windt, J. M., editors, *Open MIND*, pages 1–24. Frankfurt am Main: MIND Group.
- Seth, A. K. (2016). The real problem. Online.
- Seth, A. K. and Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B*, 371(1708):1–10.
- Seth, A. K., Suzuki, K., and Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, 2(395):1–16.
- Seth, A. K. and Tsakiris, M. (2018). Being a beast machine: The somatic basis of selfhood. *Trends in Cognitive Sciences*, 22(11):969–981.
- Shahin, A. J., Bishop, C. W., and Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *NeuroImage*, 44(3):1133–1143.
- Shahin, A. J., Kerlin, J. R., Bhat, J., and Miller, L. M. (2012). Neural restoration of degraded audiovisual speech. *NeuroImage*, 60(1):530–538.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234):303–304.
- Shapiro, L. (2019a). *Embodied cognition*. New Problems of Philosophy. London & New York: Routledge, 2nd edition.

- Shapiro, L. A. (2019b). Flesh matters: The body in cognition. *Mind & Language*, 34:3–20.
- Sims, A. (2017). The problems with prediction: The dark room problem and the scope dispute. In Metzinger, T. and Wiese, W., editors, *Philosophy and Predictive Processing*, chapter 23, pages 1–18. Frankfurt am Main: MIND Group.
- Skora, L. I., Livermore, J. J. A., Nisini, F., and Scott, R. B. (2021). Awareness is required for autonomic performance monitoring in instrumental learning: Evidence from cardiac activity. *PsyArXiv*.
- Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., and Obonsawin, M. (2004a). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness & Cognition*, 13(4):657–690.
- Smallwood, J., O'Connor, R. C., Sudberry, M. V., Haskell, C., and Ballantyne, C. (2004b). The consequences of encoding information on the maintenance of internally generated images and thoughts: The role of meaning complexes. *Consciousness & Cognition*, 13(4):789–820.
- Smallwood, J., O'Connor, R. C., Sudbery, M. V., and Obonsawin, M. (2007). Mind-wandering and dysphoria. *Cognition & Emotion*, 21(4):816–842.
- Smallwood, J. and Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, 132(6):946–958.
- Smallwood, J. and Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology*, 66:487–518.
- Smilek, D., Carriere, J. S. A., and Cheyne, J. A. (2010a). Failures of sustained attention in life, lab, and brain: Ecological validity of the task. *Neuropsychologia*, 48(9):2564–2570.
- Smilek, D., Carriere, J. S. A., and Cheyne, J. A. (2010b). Out of mind, out of sight: Eye blinking as indicator and embodiment of mind wandering. *Psychological Science*, 21(6):786–789.
- Smith, R., Kuplicki, R., Feinstein, J., Forthman, K. L., Stewart, J. L., Paulus, M. P., Tulsa 1000 investigators, and Khalsa, S. S. (2020). A bayesian computational model reveals a failure to adapt interoceptive precision estimates across depression, anxiety, eating, and substance use disorders. *PLoS Computational Biology*, 16(12):e1008484.
- Smith, R., Thayer, J. F., Khalsa, S. S., and Lane, R. D. (2017). The hierarchical basis of neurovisceral integration. *Neuroscience & Biobehavioral Reviews*, 75:274–296.

- Sohoglu, E. and Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences*, 113(12):E1747–E1756.
- Sohoglu, E. and Davis, M. H. (2020). Rapid computations of spectrotemporal prediction error support perception of degraded speech. *eLife*, 9.
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., and Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, 32(25):8443–8453.
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., and Davis, M. H. (2014). Top-down influences of written text on perceived clarity of degraded speech. *Journal of Experimental Psychology: Human Perception & Performance*, 40(1):186–199.
- Sokolov, E. N. (1960). Neural models and the orienting reflex. In Brazier, M. A. B., editor, *The central nervous system and behavior*, pages 187–276. New York: Josiah Macy Jr Foundation.
- Sokolov, E. N. (1963a). Higher nervous functions: The orienting reflex. *Annual Review of Physiology*, 25:545–580.
- Sokolov, E. N. (1963b). *Perception and the conditioned reflex*. Oxford: Pergamon Press.
- Somsen, R. J., van der Molen, M. W., Jennings, J. R., and van Beek, B. (2000). Wisconsin card sorting in adolescents: Analysis of performance, response times and heart rate. *Acta Psychologica*, 104(2):227–257.
- Sperry, R. W. (1952). Neurology and the mind-brain problem. *American Scientist*, 40(2):291–312.
- Spinoza, B. d. (2017). *The ethics (Ethica ordine geometrico demonstrata)*. Urbana, IL: Project Gutenberg.
- Spruit, I. M., Wilderjans, T. F., and van Steenbergen, H. (2018). Heart work after errors: Behavioral adjustment following error commission involves cardiac effort. *Cognitive, Affective, & Behavioral Neuroscience*, 18(2):375–388.
- Stanislaw, H. and Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1):137–149.
- Staub, B., Doignon-Camus, N., Marques-Carneiro, J. E., Bacon, E., and Bonnefond, A. (2015). Age-related differences in the use of automatic and controlled processes in a situation of sustained attention. *Neuropsychologia*, 75:607–616.

- Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A. E., Paliwal, S., Gard, T., Tittgemeyer, M., Fleming, S. M., Haker, H., Seth, A. K., and Petzschner, F. H. (2016). Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience*, 10(550):1–27.
- Sterling, P. (2020). *What is health? Allostasis and the evolution of human design*. Cambridge, MA: MIT Press.
- Stewart, J. C., France, C. R., and Suhr, J. A. (2006). The effect of cardiac cycle phase on reaction time among individuals at varying risk for hypertension. *Journal of Psychophysiology*, 20(1):1–8.
- Stiles, J. and Jernigan, T. L. (2010). The basics of brain development. *Neuropsychology Review*, 20(4):327–348.
- Strauß, A., Wöstmann, M., and Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience*, 8:350.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6):643–662.
- Struijk, P. C., Mathews, V. J., Loupas, T., Stewart, P. A., Clark, E. B., Steegers, E. A. P., and Wladimiroff, J. W. (2008). Blood pressure estimation in the human fetal descending aorta. *Ultrasound in Obstetrics & Gynecology*, 32(5):673–681.
- Suzuki, K., Garfinkel, S. N., Critchley, H. D., and Seth, A. K. (2013). Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. *Neuropsychologia*, 51(13):2909–2917.
- Teasdale, J. D., Proctor, L., Lloyd, C. A., and Baddeley, A. D. (1993). Working memory and stimulus-independent thought: Effects of memory load and presentation rate. *European Journal of Cognitive Psychology*, 5:417–433.
- Thomason, M. E. (2018). Structured spontaneity: Building circuits in the human prenatal brain. *Trends in Neurosciences*, 41(1):1–3.
- Thompson, E. (2007). *Mind in life: Biology, phenomenology and the sciences of mind*. Cambridge, MA: Harvard University Press.
- Thompson, E. and Cosmelli, D. (2011). Brain in a vat or body in a world? brainbound versus enactive views of experience. *Philosophical Topics*, 39(1):163–180.
- Thompson, E. and Varela, F. J. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences*, 5(10):418–425.

- Tort, A. B. L., Brankačk, J., and Draguhn, A. (2018). Respiration-entrained brain rhythms are global but often overlooked. *Trends in Neurosciences*, 41(4):186–197.
- Tsakiris, M. (2010). My body in the brain: A neurocognitive model of body-ownership. *Neuropsychologia*, 48(3):703–712.
- Tschantz, A., Barca, L., Maisto, D., Buckley, C. L., Seth, A. K., and Pezzulo, G. (2021). Simulating homeostatic, allostatic and goal-directed forms of interoceptive control using active inference. *bioRxiv*.
- Tuenerhoff, J. and Noppeney, U. (2016). When sentences live up to your expectations. *NeuroImage*, 124(Pt A):641–653.
- Tulving, E. and Gold, C. (1963). Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *Journal of Experimental Psychology*, 66:319–327.
- Ullsperger, M., Danielmeier, C., and Jocham, G. (2014). Neurophysiology of performance monitoring and adaptive behavior. *Physiological Reviews*, 94(1):35–79.
- Ullsperger, M. and von Cramon, D. Y. (2004). Neuroimaging of performance monitoring: Error detection and beyond. *Cortex*, 40(4-5):593–604.
- Uzzaman, S. and Joordens, S. (2011). The eyes know what you are thinking: Eye movements as an objective measure of mind wandering. *Consciousness & Cognition*, 20(4):1882–1886.
- van Boxtel, G. J., van der Molen, M. W., Jennings, J. R., and Brunia, C. H. (2001). A psychophysiological analysis of inhibitory motor control in the stop-signal paradigm. *Biological Psychology*, 58(3):229–262.
- Van de Cruys, S. (2017). Affective value in the predictive mind. In Metzinger, T. and Wiese, W., editors, *Philosophy and Predictive Processing*, chapter 24, pages 1–21. Frankfurt am Main: MIND Group.
- van der Linden, D., Keijsers, G. P. J., Eling, P., and van Schaijk, R. (2005). Work stress and attentional difficulties: An initial study on burnout and cognitive failures. *Work & Stress*, 19(1):23–36.
- van der Molen, M. W., Boomsma, D. I., Jennings, J. R., and Nieuwboer, R. T. (1989). Does the heart know what the eye sees? a cardiac/pupillometric analysis of motor preparation and response execution. *Psychophysiology*, 26(1):70–80.
- van der Molen, M. W., Somsen, R. J., Jennings, J. R., Nieuwboer, R. T., and Orlebeke, J. F. (1987). A psychophysiological investigation of cognitive-energetic relations in

- human information processing: a heart rate/additive factors approach. *Acta Psychologica*, 66(3):251–289.
- van der Molen, M. W., Somsen, R. J., and Orlebeke, J. F. (1983). Phasic heart rate responses and cardiac cycle time in auditory choice reaction time. *Biological Psychology*, 16(3-4):255–271.
- van der Molen, M. W., Somsen, R. J. M., and Orlebeke, J. F. (1985). The rhythm of the heart beat in information processing. In Ackles, P. K., Jennings, J. R., and Coles, M. G. H., editors, *Advances in psychophysiology*, volume 1, pages 1–88. Greenwich & London: JAI Press.
- van der Veen, F. M., Nieuwenhuis, S., Crone, E. A., and van der Molen, M. W. (2004a). Cardiac and electro-cortical responses to performance feedback reflect different aspects of feedback processing. In Ullsperger, M. and Falkenstein, M., editors, *Errors, conflicts, and the brain. Current opinions on performance monitoring*, pages 140–147. Leipzig: MPI for Human Cognitive & Brain Sciences.
- van der Veen, F. M., van der Molen, M. W., Crone, E. A., and Jennings, J. R. (2004b). Phasic heart rate responses to performance feedback in a time production task: Effects of information versus valence. *Biological Psychology*, 65(2):147–161.
- van der Veen, F. M., van der Molen, M. W., and Jennings, J. R. (2000). Selective inhibition is indexed by heart rate slowing. *Psychophysiology*, 37(5):607–613.
- van Es, T. (2020). Living models or life modelled? on the use of models in the free energy principle. *Adaptive Behavior*, pages 1–15.
- van Gelder, T. (1995). What might cognition be, if not computation? *Journal of Philosophy*, 92(7):345–381.
- Van Orden, G., Hollis, G., and Wallot, S. (2012). The blue-collar brain. *Frontiers in Physiology*, 3:207.
- Van Petten, C. and Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and erp components. *International Journal of Psychophysiology*, 83(2):176–190.
- van Rij, J., Wieling, M., Baayen, H. R., and van Rijn, H. (2017). itsadug: Interpreting time series and autocorrelated data using gamms.
- Varela, F. J., Thompson, E., and Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

- Vest, A. N., Da Poian, G., Li, Q., Liu, C., Nemati, S., Shah, A. J., and Clifford, G. D. (2018). An open source benchmarked toolbox for cardiovascular waveform and interval analysis. *Physiological Measures*, 39(10):105004.
- Visser, G. H., Dawes, G. S., and Redman, C. W. (1981). Numerical analysis of the normal human antenatal fetal heart rate. *British Journal of Obstetrics & Gynaecology*, 88(8):792–802.
- Walker, B. B. and Sandman, C. A. (1982). Visual evoked potentials change as heart rate and carotid pressure change. *Psychophysiology*, 19(5):520–527.
- Walls, G. L. (1954). The filling-in process. *American Journal of Optometry & Archives of American Academy of Optometry*, 31(7):329–341.
- Wang, X.-J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews*, 90(3):1195–1268.
- Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, 167(3917):392–393.
- Warren, R. M., Obusek, C. J., and Ackroff, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, 176(4039):1149–1151.
- Waselius, T., Wikgren, J., Penttonen, M., and Nokia, M. S. (2019). Breathe out and learn: Expiration-contingent stimulus presentation facilitates associative learning in trace eyeblink conditioning. *Psychophysiology*, 56(9):e13387.
- Webb, R. A. and Obrist, P. A. (1970). The physiological concomitants of reaction time performance as a function of preparatory interval and preparatory interval series. *Psychophysiology*, 6(4):389–403.
- Wessel, J. R. (2018). An adaptive orienting theory of error processing. *Psychophysiology*, 55(3):e13041.
- Wessel, J. R., Danielmeier, C., and Ullsperger, M. (2011). Error awareness revisited: Accumulation of multimodal evidence from central and autonomic nervous systems. *Journal of Cognitive Neuroscience*, 23(10):3021–3036.
- Wheatstone, C. (1838). Contributions to the physiology of vision.—part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, 128:371–394.
- Wheeler, T. and Murrills, A. (1978). Patterns of fetal heart rate during normal pregnancy. *British Journal of Obstetrics & Gynaecology*, 85(1):18–27.

- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemond, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.
- Wiese, W. and Metzinger, T. (2017). Vanilla pp for philosophers: A primer on predictive processing. In Metzinger, T. and Wiese, W., editors, *Philosophy and Predictive Processing*, chapter 1, pages 1–18. Frankfurt am Main: MIND Group.
- Wiesel, T. N. and Hubel, D. H. (1963a). Effects of visual deprivation on morphology and physiology of cells in the cat’s lateral geniculate body. *Journal of Neurophysiology*, 26:978–993.
- Wiesel, T. N. and Hubel, D. H. (1963b). Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of Neurophysiology*, 26:1003–1017.
- Wild, C. J., Davis, M. H., and Johnsrude, I. S. (2012a). Human auditory cortex is sensitive to the perceived clarity of speech. *NeuroImage*, 60(2):1490–1502.
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., and Johnsrude, I. S. (2012b). Effortful listening: The processing of degraded speech depends critically on attention. *Journal of Neuroscience*, 32(40):14010–14021.
- Wilke, C. O. (2020). cowplot: Streamlined plot theme and plot annotations for ‘ggplot2’.
- Wilkinson, M., McIntyre, D., and Edwards, L. (2013). Electrocutaneous pain thresholds are higher during systole than diastole. *Biological Psychology*, 94(1):71–73.
- Williams, D. (2018a). Pragmatism and the predictive mind. *Phenomenology & the Cognitive Sciences*.
- Williams, D. (2018b). Predictive processing and the representation wars. *Minds & Machines*, 28(1):141–172.
- Williams, D. (2020). Is the brain an organ for prediction error minimization? *Preprint*.
- Wilsch, A. and Obleser, J. (2016). What works in auditory working memory? a neural oscillations perspective. *Brain Research*, 1640(Pt B):193–207.
- Wilson, A. D. and Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology*, 4(58):1–13.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625–636.

- Wilson, R. A. (1994). Wide computationalism. *Mind*, 103(411):351–372.
- Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 65:95–114.
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 73(1):3–36.
- Wood, S. N. (2017). *Generalized additive models: An introduction with R*. Texts in Statistical Science. Boca Raton, FL.: CRC Press, 2nd edition.
- Wöstmann, M., Herrmann, B., Maess, B., and Obleser, J. (2016). Spatiotemporal dynamics of auditory attention synchronize with speech. *Proceedings of the National Academy of Sciences of the United States of America*, 113(14):3873–3878.
- Wöstmann, M., Herrmann, B., Wilsch, A., and Obleser, J. (2015). Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. *Journal of Neuroscience*, 35(4):1458–1467.
- Wöstmann, M., Lim, S.-J., and Obleser, J. (2017). The human neural alpha response to speech is a proxy of attentional control. *Cerebral Cortex*, 27(6):3307–3317.
- Wozniak, M. (2019). How to grow a self: Development of the self in a bayesian brain. *Preprint*.
- Yang, X., Jennings, J. R., and Friedman, B. H. (2017). Exteroceptive stimuli override interoceptive state in reaction time control. *Psychophysiology*, 54(12):1940–1950.
- Yon, D., de Lange, F. P., and Press, C. (2019). The predictive brain as a stubborn scientist. *Trends in Cognitive Sciences*, 23(1):6–8.
- Yuste, R. (1997). Introduction: Spontaneous activity in the developing central nervous system. *Seminars in Cell & Developmental Biology*, 8(1):1–4.
- Yutzey, K. E. and Kirby, M. L. (2002). Wherefore heart thou? embryonic origins of cardiogenic mesoderm. *Developmental Dynamics*, 223(3):307–320.
- Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., and Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *NeuroImage*, 32(4):1826–1836.
- Zelano, C., Jiang, H., Zhou, G., Arora, N., Schuele, S., Rosenow, J., and Gottfried, J. A. (2016). Nasal respiration entrains human limbic oscillations and modulates cognitive function. *Journal of Neuroscience*, 36(49):12448–12467.

- Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., Poeppel, D., and Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron*, 77(5):980–991.
- Zoefel, B. and VanRullen, R. (2015). Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *Journal of Neuroscience*, 35(5):1954–1964.