Have we lost our minds?
An approach to multiscale dynamics in the cognitive sciences

Maxwell James Ramstead
Department of Philosophy and
Division of Social and Transcultural Psychiatry
Department of Psychiatry
McGill University, Montreal

A thesis submitted to McGill University
in partial fulfilment of the requirements of
the degree of Doctor of Philosophy in Philosophy.

Table of contents

1. Abstract

1.1. Abstract (English)

This doctoral thesis examines the puzzle of how best to study multiscale systems in the cognitive sciences; namely, systemic dynamics that span several spatial and temporal scales. The aim of this thesis is to articulate – and to work out the philosophical implications of – a first principle approach to multidisciplinary research in the cognitive sciences. The aim is to model the dynamics of cognitive systems at, and across, all the spatial and temporal scales at which they exist – from cells to societies, and everything in between. To achieve this aim, this thesis proposes three novel theoretical pillars: (1) *variational neuroethology*, a new approach to neuroethology – the study of adaptive behaviour and its control – based on a general account of the information-theoretic constraints on cognizing organisms, called the free-energy principle; (2) *variational ecology*, an approach to cognitive ecology that provides a more specific account of how hierarchically structured organisms – that construct their own ecological niches – align themselves with their niche; and (3) the *cultural affordances framework*, a still more specific account of how human action, cognition, learning, and culture are organized in environmental affordances and corresponding (cultural) regimes of attention. The thesis develops a philosophical argument, based on these frameworks, *against essentialism* about the boundaries of cognitive systems in the philosophy of the cognitive sciences (frameworks that privilege one type of boundary over others). The thesis argues that the boundaries of cognitive systems are *multiple*, *nested*, and *interest-relative* – and provides a promising resolution of the dialectic between internalism and externalism in the philosophy of the cognitive sciences.

1.2. Abstract (French)

Cette thèse cherche à résoudre une énigme, à savoir : quelle est la meilleure manière d'étudier les systèmes multi-échelles – les dynamiques systémiques qui s'étendent sur plusieurs échelles spatiales et temporelles. L'objectif de cette thèse est d'articuler une approche multidisciplinaire pour les sciences cognitives qui puisse se fonder sur des principes premiers – et d'en distiller les conséquences philosophiques. L'objectif est de permettre la modélisation des dynamiques des systèmes cognitifs à (et à travers) toutes les échelles auxquelles ces systèmes existent – des cellules aux sociétés, en passant par tout le reste. Pour réaliser cet objectif, cette thèse propose trois nouveaux piliers théoriques : (1) la *neuro-éthologie variationnelle*, qui est une nouvelle façon d'aborder la neuro-éthologie – c'est-à-

dire l'étude du comportement adaptatif et de son contrôle – basée sur une formulation générale des contraintes auxquelles doivent se plier tout système cognitif, provenant de la théorie de l'information (le principe de minimisation de l'énergie libre) ; (2) l'écologie *variationnelle*, qui est une nouvelle perspective (plus spécifique) ouverte sur le projet des écologies cognitives, dont le but est d'expliquer les mécanismes qui permettent l'alignement mutuel de groupes d'organismes dotés d'une structure hiérarchique et qui construisent ensemble leur niche écologique, d'une part, et de leur niche partagée, d'autre part ; enfin, (3) la théorie des affordances culturelles, un cadre théorique (encore plus spécifique) qui rend compte de l'organisation de la culture, de l'action, de la cognition, de l'apprentissage humains, par le biais des affordances environnementales et des régimes de l'attention (culturels) correspondants. Cette thèse développe un argument, basé sur ces cadres théoriques, *contre l'essentialisme* par rapport aux frontières des systèmes cognitifs dans la philosophie des sciences cognitives (c'est-à-dire, les théories qui privilégient un seul type de frontière). Cette thèse fournit l'argument selon lequel les frontières des systèmes cognitifs sont *multiples*, *enchâssées*, et *relatives aux intérêts des chercheurs* – et propose une résolution de la dialectique entre l'internalisme et l'externalisme dans la philosophie des sciences cognitives.

2. Acknowledgements

Dedicated to my grandfather, Bernard Désormeau,
and to my grandmother, Françoise Désormeau Couture.
Grand-papa, tu m'as toujours dit :
« Si tu fais quelque chose, fais-le bien, ou bien ne le fais pas du tout. »
J'espère que tu trouveras que j'ai bien fait.

This thesis would not have been possible without the help and support of several people, groups, and institutions. The first group of people that I wish to thank comprises my mentors.

To begin, I want to express my deep and unending gratitude and admiration to Laurence Kirmayer. Laurence has been with me from the start of this adventure, taking me on when I was a young Master's student. Laurence has become a father figure and, dare I say, a dear friend. Laurence, it may sound cheesy, but you have inspired me to achieve my dreams. You've shown me what it means to work as a team. Thanks to you, I found an intellectual home at the Division of Social and Transcultural Psychiatry and a group of friends and peers at the Culture and Mental Health Research Unit. I am, and I will always be, grateful to you.

Karl Friston is an intellectual titan. I have rarely been starstruck, but meeting Karl in May 2016 was one of those times. It was my pleasure to learn first-hand that he is also a kind, endearing, and witty soul. Karl, you took me under your wing – not only showing me the ropes, but becoming one of my PhD supervisors and friends as well. I have learned *so much* from you. I am deeply grateful for the warm welcoming atmosphere that you provide us at the Functional Imaging Lab and for your constant, friendly support and mentorship. Thanks to you, London has become my second home.

Ian Gold deserves thanks for being one of the kindest humans that I have ever met in my life. Ian, you were there for me when things got crazy and you were a ray of light when I needed it most, and you have helped to guide me through this degree. I am deeply grateful for your help and support.

Samuel Veissière, you are my brother. Sam, I love you. You showed me what it means to be an adult in academia. What has resulted from our collaboration has been an experience of genuine intellectual connection and mutual recognition, which grew into one of the most rewarding friendships that I have ever experienced. I can say without the shadow of a doubt that one of the greatest experiences in my entire life was our road trip/conference travel to Australia. This was an initiation rite of the deepest significance and the experience

has stayed with me. In my career as a graduate student, I could not have hoped to be supported as consistently and generously by a mentor and friend.

Pierre Poirier, even after all these years, you still find ways to support me. At the risk of repeating myself, I could not have asked for a better Master's supervisor. Your constant interest, care, compassion, and upbeat nature have made this adventure all the better. I will always be grateful for your helping me find my way.

The next person to thank is my very close friend and collaborator, Axel Constant. Axel, you are just the best. No, seriously. You are a star. You are kind, fun, brilliant, and reliable – the perfect work partner. I look forward to a sunny future working alongside you. You were my first student, and I have had the privilege of seeing you blossom into a full-fledged, amazingly talented, insanely creative genius. I love you bro.

The next person I want to thank is Paul Badcock – from the bottom of my heart. Paul, you're such a lovely, loving human. I have the fondest memories of discovering new places with you. Love you brother. You were the first person I met through my collaboration with Karl, and our friendship-collaboration set the tone for years to follow – and set a high bar. You're such a great guy and an inspiring (not to mention inspired) scholar. I'm deeply enthusiastic about our continuing collaboration and I'm so, so proud of what we've already accomplished together. Shoulder bump and hugs.

Next to thank is Casper Hesp. Casper, you are just amazing. I have learned so much from you. You're my brother from far away. You made my experience at the Functional Imaging Lab such an exciting and life changing experience – the Dutch and Canadian duo. I'll always be grateful for your kindly showing me around London! Genuinely I'm excited to have such a cool and gifted close friend and collaborator. I truly look forward to our future collaborations and projects.

Next, I would like to thank collaborators who have become friends and mentors: Michael Kirchhoff and Micah Allen. I met Micah and Michael through the internet – a very postmodern way to start a friendship. Michael and Micah, I find the two of you inspiring. I hope to strike the kind of work-life balance that you have achieved, and to have as love-filled a personal life as you both. You've shown me that Hume's advice still rings true: "Be a philosopher; but, amidst all your philosophy, be still a man." I have tried very hard to emulate you – I hope to live in a house in the bush, with my partner and our dogs, and my log cabin (like Michael and Heidegger).

Next, I would like to thank my long-standing besties: Gustave Lavoie Prud'homme, William-J. Beauchemin, Nabil Bouizegarene, Joel Entwistle, and Sébastien Coté. I love you

guys so much. Thank you for being there – I wouldn't be here without you. Thanks also to my friends who have helped me grow, personally and intellectually: Ian Robertson, Jonathan St-Onge, Safae Essafi Tremblay, Simon C. Tremblay, Thomas Parr, David Benrimoh, Nick Brancazio, Auguste Nahas, Eric White, Michael Lifshitz, Ana Gómez-Carrillo De Castro, Iris Rapoport, Jared Vasil, Mel Andrews, Liza Solomonova, Bruno Dubuc, Ishan Walpola, Ceren Kaypak, Patrick Garon-Sayegh, Thomas Thiery, Yann Harel, Adam Safron, Josh Bamford, Viky Neascu, Jacob Taylor, Moriah Stendel, Aron Vallinder, Ariela Lavana, Alison Starkey, Naila Kuhlmann, Marcelle Partouche Gutierrez, Frank Muttenzer, Philippe Blouin, Ariane L'heureux-Garcia, Vincent Laliberté, and Ryan Smith. Thanks also to my students, who are deeply inspiring and will do great things: Alice Hickling and Irene Arnaldo.

Obviously, I am deeply grateful to all those who read previous drafts of my thesis: my doctoral supervisors Laurence Kirmayer, Ian Gold, Karl Friston, as well as Axel Constant, Pierre Poirier Samuel Veissière, Safae Essafi Tremblay, Simon C. Tremblay, Adam Safron, Ian Robertson, and Nick Brancazio.

Finally, I want to give special thanks to everyone who supported me in my personal life: Maman, Babs, Laurie, Dad, Jack, Gloria – I love you all deeply. William-J. Beauchemin, you have been a daily source of support and you are the best flatmate a guy could ask for. Thanks to my other half, Zoë, for being there, and for being loving, open, compassionate, and understanding. I love you. Finally, merci grand-papa and grand-maman. Je vous aime.

To finish, I would like to express my deep and sincere gratitude to my funders. They made this research possible in a very real, material sense. My doctoral degree was funded by a Joseph-Armand Bombardier Canada Graduate Scholarship courtesy of the Social Sciences and Humanities Research Council of Canada. I also benefitted from a Michael Smith Foreign Study Supplements from the same organisation, which allowed me to undertake one of the most amazing adventures of my life: my first internship at the Wellcome Trust Centre for Neuroimaging's Functional Imaging Lab, at University College London with Karl Friston's group. I would also like to thank deeply the current provider of my funding, the Canada First Research Excellence Fund, which was awarded to McGill University for the Healthy Brains for Healthy Lives initiative, and which in turn funded my last year of doctoral work, as well as the Office for Graduate and Postgraduate Studies at McGill, which awarded me with a Graduate Mobility Award to support my second internship in London.

Thank you all so much.

Love,

Maxwell

3. Copyright notice and references to original publications

I gratefully acknowledge the support provided by the Wellcome Trust Centre for Neuroimaging at University College London and the Division of Social and Transcultural Psychiatry at McGill University for providing open access to all the articles collected in this thesis. Thanks to their support, I am the copyright holder for these articles and I exercise my right to include them as part of my doctoral dissertation.

The first chapter, entitled "Cultural affordances: Scaffolding local worlds through shared intentionality and regimes of attention," was first published as an original research article in *Frontiers in Psychology* (Ramstead, Veissière, & Kirmayer, 2016). This article is distributed under the terms of the Creative Commons Attribution License (CC BY).

The second chapter, entitled "Answering Schrödinger's question: A free-energy formulation," was originally published as a review paper in the journal *Physics of Life Reviews* (Ramstead, Badcock, & Friston, 2018a). This article is distributed under the terms of the Creative Commons Attribution License (CC BY 4.0). The paper received target article treatment with 14 peer commentaries, also at *Physics of Life Reviews* (Allen, 2018; Bellomo & Elaiw, 2018; Bitbol & Gallagher, 2018; Bruineberg & Hesp, 2018; Campbell, 2018; Daunizeau, 2018; Kirchhoff, 2018; Kirmayer, 2018; Leydesdorff, 2018; Martyushev, 2018; Pezzulo & Levin, 2018; Tozzi & Peters, 2018; Van de Cruys, 2018; Veissière, 2018). We were invited to write a response by the journal as well (Ramstead, Badcock, & Friston, 2018b).

The third paper, "Variational ecology and the physics of sentient systems," was originally published as a review paper in the journal *Physics of Life Reviews* (Ramstead, Constant, Badcock, & Friston, 2019). This article is distributed under the terms of the Creative Commons Attribution License (CC BY 4.0).

The fourth and final chapter of this thesis, "Multiscale integration: Beyond internalism and externalism," was originally published as a research article in the journal *Synthese* (Ramstead, Kirchhoff, Constant, & Friston, 2019). This article is distributed under the terms of the Creative Commons Attribution License (CC BY 4.0).

4. Introduction

4.1. Have we lost our minds?

The questions raised by the *integrated, adaptive behaviour* of *multiscale systems* are at the heart of this doctoral thesis. Generally speaking, the cognitive sciences study the adaptive behaviour and cognitive processes of organisms, which are living systems that exhibit intelligent, situationally appropriate forms of action. The crucial observation that underwrites my doctoral work is that most of these systems are systems of systems. Multicellular organisms are composed of systems embedded in larger systems, which are, in turn, embedded in still larger systems: cells are embedded in organs, which interact in a coordinated way to give rise to the activities of organisms, which are themselves embedded in larger social groups. Each such system, at every scale, interprets its respective environment in order to coordinate meaningful patterns of behaviour. Somehow, by some apparent miracle, the whole system of systems moves together in an integrated fashion. How does this differentiation and integration arise?

Consider a human sociocultural group; say, the McGill University community in Montreal, Canada. Any given human society is composed of several, densely coupled components; some of these are living, thinking agents, and others are material artefacts, codes of conduct, norms, places, and practices, which together constitute an ecological niche. In the case of McGill, the former group of components comprises undergraduate and graduate students, postdoctoral fellows, professors, lecturers, research staff, other faculty members, administration and support staff, and so on; the latter comprises large parts of Mount Royal, as well as the buildings where teaching and research take place, not to mention lab equipment and research materials. The denizens of McGill engage in those patterned cultural practices that define McGill University as a sociocultural community, e.g., lecturing and listening, note taking, researching, supervising, form filling, thesis writing, and so on. The agents that make up the McGill community, in turn, are themselves composed of systems that are also alive, and which must perform situationally appropriate actions at a subordinate scale. The student body, thus, is composed of human bodies, themselves made up of organs, which function thanks to the regulated operation of their component cells. Tissues and cells (even when cancerous) also interpret their environment (human bodies) to act in situationally appropriate ways (appropriate for the cancer, if not the host). These systems are nested into one another along increasing spatial scales.

Somehow, all these moving parts are integrated in – and indeed, through – adaptive behaviour, in loops of action and perception. The coordinated, well-regulated action of cells allows for the good functioning of organ tissues, such as the heart and brain; the coordinated integration of the functions performed by myriad organs of the human body results in perception and action by human agents, both automatic and intentional; and agents themselves interact in a way robust and coherent enough for us to speak of, say, McGill University as a distinct sociocultural community, with its own values, norms, and ways of doing things.

To explain the coherence of systems as complex as human communities and their dynamics, the way they act, their manner of changing and evolving time, clearly appeals to processes that unfold over several scales. We have just considered spatially nested systems, in which small component systems are progressively embedded into larger ones. Something similar can be said about temporal scales. At real-time or mechanistic timescales – which comprise events on the order of milliseconds to seconds and minutes – we need to explain how organisms are able to perceive, interpret, and act upon their worlds in situationally appropriate ways; at ontogenetic or developmental timescales – on the order of minutes to years – we need to explain how the bodies and brains of organisms mature and develop; for *Homo sapiens*, this means to explain how human agents become encultured and learn the norms that govern their societies, i.e., 'the kind of thing that we do here'; finally, at evolutionary and phylogenetic timescales – events unfolding over intergenerational timescales – we need to explain the deep evolutionary and historical processes that have led to modern human phenotypes. Clearly, this sequence of scales also evinces a nested structure. And by some miracle of nature, all these processes unfold in an integrated way.

All of these processes contribute meaningfully to what we call 'mind', and in ignoring the amazing complexity of this nested structure of *systems of systems* raises a danger – that of *losing our minds*.

In ignoring the nested structure of cognitive systems, we tun the risk turning a blind eye to some of the most important phenomena for understanding and explaining adaptive behaviour and action. This raises deep questions for the cognitive sciences: What do the cognitive sciences look like when the issue of multiscale dynamics takes centre stage? What do explanations look like in such a regime? How does systemic integration take place across scales? How can we best study multiscale systems, at and across all the scales at which they exist – from cells to societies?

This thesis explores theoretical and methodological questions related to the study of multiscale cognitive systems in the cognitive sciences. In other words, the thesis clarifies what is at stake in the modelling of systemic dynamics that span several spatial and temporal scales. The aim of this doctoral thesis is to articulate – and examine the philosophical implications of – a framework for research in the cognitive sciences to study the multiscale systemic dynamics of the mind.

4.2. Previous attempts to address multiscale systemic dynamics in philosophy

That phenomena of mind involve systems that exist at and across multiple different spatial and temporal scales has, of course, not gone unnoticed – and has stoked lively debate in philosophy in the cognitive sciences. Broadly speaking, this topic has been taken up and explored in the context different theoretical debates, the two most well-known of which are that between *reductionist* and *emergentist* approaches to the ontology and epistemology of cognitive systems, and that between *internalist* and *externalist* approaches to the boundaries of cognition. Here, I briefly discuss the first project and explain why I have chosen largely to avoid casting the issue of multiscale dynamics in the terms posed by this debate. I then motivate my approach by discussing the dialectic at play in a debate more aligned with my aims.

4.2.1. Reductionism and emergentism

The idea that the systems studied by the natural sciences are structure as nested systems of systems is not new; contemporary philosophical work on this topic goes back at least to the British emergentists (Ablowitz, 1939; Broad, 2014/1925; Mill, 1843; Morgan, 2013/1923); for a historical review, see (McLaughlin, 1992). This literature emphasised the idea that natural systems in general were endowed with a hierarchical structure: atoms form molecules, which are the constituent parts of cells, which make up organs – and so on. ‘Higher-order’ phenomena were said to emerge from ‘lower-order’ ones. Debates abounded over how to understand the hierarchical structure and what exactly were the levels at play – the history of which is far beyond the scope of this doctoral thesis.

Between the 1970s and 2000s, the idea of emergence was mainstreamed via its central place in the debate over the epistemological and ontological status of the causal powers of emergent phenomena. This debate centred on the idea of *downward causation* (Campbell, 1974). This view is premised on the idea that biological systems are hierarchically organised, that is, that living systems evince nested levels of organisation. Downward causation is the

view that phenomena occurring at higher levels of organisation have causal effects on the world and on their constituent parts – causal effects that were not merely derivative from, or epiphenomena of, the causal powers of their constituent parts.

The debate in that raged philosophy over emergentism has tended to focus on whether it makes sense to speak of downward causation. The two main flavours of emergentism are ontological (or strong) emergentism and epistemological (or weak) emergentism. Strong emergentism is the view that emergent phenomena have independent existence and causal powers, ontologically speaking, from the phenomena from which they emerge (Bishop, 2008; Boogerd, Bruggeman, Richardson, Stephan, & Westerhoff, 2005; Juarrero, 1999; Kauffman, 1993; Maturana & Varela, 1980; Thompson, 2010). This view has come under much critical scrutiny; for instance, some argue that to speak of downward *causation* conflates relations of causation (which can only hold between two distinct things) and those of *composition* (Bechtel, 2009; Kistler, 2009). The most popular position seems to be that the causal powers of emergent phenomena are inexistent; emergent phenomena are epiphenomenal and downward causation is not philosophically respectable (Kim, 1999, 2006).

Weak emergentism, which today seems to be the dominant version of emergentism (Bedau & Humphreys, 2008), is the view that emergence is an *appearance* that is an effect of the limitations of our scientific (or broader explanatory) models – i.e., the view that whether a set of phenomena are emergent or not is determinable only relative to some theory that can predict the properties of emergent phenomena (Bedau, 1997, 2002; Craver, 2007; Craver & Bechtel, 2007; Hempel & Oppenheim, 1948). The absence of strong metaphysical commitments typically favours this flavour of emergentism.

From my point of view, the debate over reductionism and emergentism debate is not the most productive, nor particularly generative, to answer questions about how best to study multiscale systems. The fact that biological systems possess a clear hierarchical organisation is no longer debated in biology and the cognitive sciences. It is an established, measurable fact about biological systems that they evince such a structure. The causal role of hierarchical structure, for example, in increasing information processing efficiency in hierarchically organised nervous systems, or increasing robustness to environmental perturbation, is increasingly well understood (Perdikis, Huys, & Jirsa, 2011; Sporns, 2013). Indeed, such a scientific understanding of the causal effects of hierarchical nesting predates the debate in philosophy by some time (Simon, 1965). Settling metaphysical debates about the causal powers of emergent phenomena does little to help us answer questions related to how best to study multiscale systems.

Some in the cognitive sciences have proposed a fruitful resolution of this dialectic by suggesting that we must be both emergentists *and* reductionists when we study complex nested systems (Bechtel, 2007, 2009; Bechtel & Abrahamsen, 2010). That we must be emergentists means that we must recognise that emergent phenomena exist, that they seem to be causally relevant – they seem to have real causal effects on the world in which they are embedded *qua* emergent phenomena. That we must be emergentists means that we must engage with the complexities of systems that exist at several scales. That we must also be reductionists means, in a complementary fashion, that we still ought to strive to produce mechanistic explanations for such emergent phenomena – speaking to the ways in which higher-order system dynamics organise constituent parts of the system, and the ways in which the system affects its embedding environment *qua* system. Indeed, it may be that focusing on downward causation, as the debate between emergentists and reductionists has tended to do – debates over the metaphysical status of the effects of system-level dynamics on the components of the system – misses some of the crucial ways in which emergence displays its causal efficacy; namely, through the dynamics of the whole system and through its effects, as a system, on its embedding environment.

From a mathematical point of view, formal models exist that can accommodate either of these perspectives. (Non-technically minded readers are invited to skip this brief mathematical excursus.) The theoretical dialectic here is that between *synergetics* and *renormalization group theory*. Synergetics provides a formal framework for modelling circular causality, i.e., for modelling bidirectional causal relations between dynamics unfolding at different levels of description (Haken, 1977; Haken, 1983, 2013). In the synergetics approach, the microstates that compose a lower-order phenomenon are averaged to yield macrostates, which form phenomena at the higher scale. In this sense, mixtures of microstates cause macrostates. The other direction is ensured via the enslaving principle. Briefly, by construction, macrostates evolve more slowly than microstates. Technically, this is because decay back to average dynamics of the entire system (aka, the centre manifold) is faster than the flow on the centre manifold itself. In the synergetics model, it will look as if the centre manifold is enslaving all local perturbations away from that manifold, which licences us to speak of a causal influence of higher-order or emergent phenomena (macrostates) on the systems from which they emerge (microstates).

Renormalization group theory, which originated in quantum mechanics (Shirkov, Bogoliubov, & Bogoliubov, 1959), provides the opposite perspective on emergent phenomena; i.e., that causality flows only one way, from lower- to higher-order phenomena.

The basic idea is that if there is some way of describing a system and that there is some way of moving from descriptions at one scale to the next, such that the form of any complete (dynamical) description of system remains unchanged, then that system is said to be *renormalizable*. One perplexing feature about such mathematical systems is that such transformations are often unidirectional. That is, while one can move from microstate descriptions to macrostate descriptions, one cannot recover the microstates that went into the mix to yield the macrostates. This asymmetry means that in renormalizable systems, microscopic states cause macroscopic states in a strong sense, and that we cannot speak of deterministic relations in the inverse direction.

From my point of view, new developments in the cognitive sciences and mathematics render the debate between emergentists and reductionists moot – especially constructs and techniques from systems theory, which permeate and guide much of the work in this thesis. (Von Bertalanffy, 1950, 1972). Systems theory is an entire field of enquiry devoted to the study of system-level dynamics and processes. Such mathematical techniques have enabled scientists working from a systems theoretic perspective to study systemic dynamics rigorously – leading to the emergence of 'systems thinking' (Weinberg, 1975) and indeed, to entirely new fields of enquiry organised around the study of system-level causal effects and systemic dynamics, including computational systems biology (Kitano, 2002), systems neuroscience (Van Hemmen & Sejnowski, 2005), and systems ecology (Odum, 1994).

In any event, questions about whether or not emergent phenomena have causal effects that are independent of those of the systems from which they emerge, and if so to which extent they are, seem like *empirical* questions. Settling questions about the metaphysics of downward causation in emergent systems tells us little about how best to *study* their effects, *qua* integrated system, on its embedding environment. More importantly, it tells us nothing at all about how the dynamics of such systems are *integrated across scales* – that is, how the system can act on the world in an integrated fashion despite having constituents at different scales.

One might object to this as follows: if we were able to establish that emergent properties do (or do not) have top-down causal effects or causal powers independent from those of their realisation base, this would in effect provide constraints on the kinds of allowable explanations and might guide interpretations of the phenomenon to be understood. This, however, begs the question of *how* one might establish the fact that emergent properties have such causal powers at all. It seems like there is no possible recourse to formal theories

(e.g., to mathematical models) to settle these issues – since both the emergentist and the reductionist options are formally defensible.

Whether or not some phenomenon should be studied as an emergent phenomenon seems to depend on the phenomenon being studied – and mathematical tools exist that can be used to develop and justify either approach. It may be, for instance, that emergentist models based on the synergetic approach are better suited to explain biological or cognitive phenomena; and that renormalization group theoretical approaches are, in turn, better suited to approach dynamics at quantum scales. We return to these issues surrounding emergentism and reductionism – and provide a sustained argument for our rejection of this debate – in the ninth section of this thesis, which provides a comprehensive discussion of my findings.

4.2.2. Internalism and externalism

For my purposes, the debate in the philosophy of the cognitive sciences between *internalists* and *externalists* about the *boundaries of cognitive systems* is more prescient. This debate is over where to draw the boundaries of cognitive systems. In the cognitive sciences, there has been progress in at least trying to deal with the complexities of systems of systems, which often explicitly draw upon the mathematical resources just discussed from complex systems theory – making this a more fruitful point of departure.

It is worth mentioning that, prior to (and into) the 1980s, many actors in the cognitive sciences were committed explicitly to the idea that the workings of the brain were only relevant as 'mere implementation' – mere plumbing, in effect – and that implementation had little or nothing to say about cognitive processes and the mind. Cognitive scientist David Marr, for instance, proposed an epoch-defining framework to study cognition that feature three distinct levels of explanation: the computational level, which specifies the information-processing problem to be solved; the algorithmic level, which specifies algorithms and representations that are capable of solving the problem specified at the computational level; and the implementation level, which specifies the physical structures that realise these algorithms and representations (Marr, 1982). Intelligence was seen as relating to the algorithms and representations that were involved in cognitive processes – with implementation seen as mere machinery.

Since the 1980s, however, in part thanks to the rise of sophisticated experimental methods like genetic sequencing and neuro-imaging, and the shift towards biological classificatory systems in psychiatry (e.g., the third edition of the Diagnostic and Statistical Manual of Mental Disorders and the Research Domain Criteria framework proposed by the

US National Institute of Mental Health), the main models of explanation in the cognitive sciences have, above all, tended to have an internalist, or neuro-reductionist, flavour (Bechtel, 2007; Bracken et al., 2013; Choudhury & Slaby, 2016; Gold, 2009; Gold & Kirmayer, 2007; Guze, 1989; Kirmayer & Gold, 2012). The main characteristic of these internalist models of explanation is that they identify the only relevant causal determinants of human behaviour and mental life as being those internal to the boundary of the skull (e.g.., neural processes) or the boundary of the organism (e.g., genetic and epigenetic processes). Of course, any study of the brain or genome will feature phenomena that 'point outwards' (the brain represents features of the outside world, the genome and especially epigenome are reactive to environmental states and situations), but on these views, the causally relevant machinery is still housed within the boundary of skull or skin.

Psychiatry provides a good case study of internalism. Today, psychiatry is understood by those who share the internalist view as "clinically applied neuroscience" (Insel & Quirion, 2005), that is, as a science the scope of which covers internal, biological causal determinants of human behaviour and mental life (especially those located in the brain). The mainstreaming of the internalist model of explanation has, accordingly, tended to privilege methods investigating causal factors internal to the organism; for example, genetic sequencing and animal model studies, single-cell neural recordings, structural/functional neuroimaging, and psychopharmacology – none of which foreground factors outside the organism. The prominence of internalist model in the cognitive sciences has – to the chagrin of many, and despite criticism (Choudhury & Slaby, 2016) – side-lined approaches that instead focus on the causal determinants that are operative at other spatial and temporal scales – such as the influence of causal factors such as cultural group identification, first-person phenomenological experience, socioeconomic status (Kirmayer & Ramstead, 2017; Kirmayer, Lemelson, & Cummings, 2015). Of course, we should note that the cognitive sciences are broader in scope that their applications to psychiatry – and that psychiatry raises additional issues related to pathology and context that may be less immediately apparent in some cognitive sciences. But this is the general picture: for having too narrowly focused on internal factors, psychiatry has lost sight of much of what plays a role in the working of human minds.

The problems and limitations of internalism did not go unnoticed. By now, there is a well-established and still growing *externalist* literature that has pushed back against internalist views. Externalism, broadly, is the view that factors outside the boundary of the skull or skin can meaningfully be considered to be involved in cognitive processes, and that

the boundaries of cognitive systems spill out or cross into the external world (Gallagher, 2017; Hutto & Myin, 2013; Thompson, 2010).

Externalist views are, on the face of it, quite amenable to act as the theoretical foundation of an approach to multiscale systems and their dynamics. Their aim is to study cognitive systems in a way that does justice to the 'external' causal determinants of mental life – those factors that play causal roles in the generation of adaptive behaviour, but which lie beyond the boundaries of skull and skin (Kirmayer et al., 2015; Newen, De Bruin, & Gallagher, 2018). Many authors today emphasise 'external' causal processes, including: the role of the body in cognition, a position known as embodied cognition (Gallagher, 2006; Noë, 2004; Varela, Thompson, & Rosch, 1991); the role of action in cognition and perception, known as the enactive approach to cognition (Engel, Friston, & Kragic, 2016; Gallagher, 2017; Thompson, 2010); and the role of context and broader cognitive ecologies in the generation by organisms of adaptive action, variously defended by proponents of extended and embedded cognition (Clark, 2008) and by advocates of ecological psychology (Gibson, 1979; Hutchins, 2010) – all of which we review in detail below.

Unfortunately, this shift to externalist approaches led to scientists and philosophers throwing out the metaphorical baby with the metaphorical bathwater. Externalists typically rejected the tools and metaphors that drove internalist approaches, such as the digital computer metaphor of the brain, favouring more dynamical metaphors like the brain as a steering wheel or a Watt governor (Van Gelder, 1998); or the idea that the brain or genes encode information (Thompson, 2010). Thus, the rejection of internalism led to the rejection of the tools, constructs, and methods typically associated with internalism, namely, information theory and computational modelling.

This seems to have become the dominant view in externalist theories by the early 2000s. The enactive philosophers Shaun Gallagher and Alva Noë, for instance, reject computational and information-theoretic explanations in the cognitive sciences (Gallagher, 2006, 2017; Noë, 2004). Philosopher Evan Thompson, in his magnum opus *Mind and Life*, which has become the central text of enactivism, rejects the use of information theory and notions such as encoding as being anything but useful metaphors in the cognitive sciences (Thompson, 2010). The most extreme form of this position is the radical enactivism of philosophers Dan Hutto and Erik Myin. In a nutshell, radical enactivism rejects the appeal to internal representations to explain adaptive behaviour. Radical enactivists are suspicious of appeals to information theory and computation to explain cognitive processes, and more precisely, believe that we will not explain cognitive *semantics* (i.e., the representational

contents of experience) by appealing to constructs and strategies from information theory (Hutto & Myin, 2017; Hutto & Myin, 2013). We review these approaches, as well as what we take to be their limitations, below.

Part of the original work accomplished by this thesis is to reconcile the *tools* that power more internalist approaches to the study of cognitive systems with a broadly pragmatist or enactive view; and indeed, to rehabilitate the ideas that animated internalist approaches in a way that makes them amenable and complementary to externalist approaches. A multiscale approach cannot merely consist in the rejection or denial of internalist approaches.

4.3. The approach presented here

This thesis develops three new theoretical frameworks that collectively function as a new model of explanation for the cognitive sciences. A model of explanation is a theoretical model of the underlying assumptions operative in a given science, the function of which is to account for how that science is able to construct explanations for the phenomena within its explanatory scope (Bechtel, 2007; Craver, 2007; Cummins, 1983; Kendler & Parnas, 2008; Murphy, 2010; Schaffner, 1993). There are many such models in the different sciences, each appealing to different types or families of causal factors to generate explanations of the phenomena within their purview. This doctoral thesis focuses on models of explanation that are operative in the cognitive sciences – which I use as an umbrella term to covers the scientific, multidisciplinary study of the mind and behaviour, including philosophy, psychology, neuroscience, psychiatry, biology, ethology, anthropology, and computer science (among other disciplines).

This thesis introduces a new reading of – and an argument against – *essentialism* in the context of debates in the philosophy of the cognitive sciences over where to draw the boundaries of cognitive systems. This usage is not typical in philosophy and should be justified briefly. Essentialism in philosophy typically relates to the use of necessary and sufficient conditions for inclusion in a given category. With regard to the issue of where to draw the boundaries of cognitive systems, essentialism – as defined in this thesis – is the view according to which cognitive systems must be defined relative to one privileged type of boundary. Essentialism is the main critical target of this thesis. Indeed, in the cognitive sciences, most models of explanation are essentialist. The four chapters of this thesis build upon each other, culminating in the final chapter, which provides an argument against essentialist views on the boundaries of cognitive systems in the cognitive sciences. Against

essentialism, this thesis proposes ontological and methodological *pluralism*. Ontological pluralism means that the boundaries of cognitive systems are *multiple*, *nested*, and *interest-relative*. This means that any definition of cognitive systems in terms of necessary and sufficient conditions, relative to any single type of boundary, is *arbitrary* and *mistaken*. Methodological pluralism follows from the ontological variety, and entails that a variety of different methods, operating at different scales and inspecting different boundaries, is necessary to study the mind; this, in turn, entails abandoning the idea that there is one single method that will be appropriate to study cognitive systems.

Today, the most common form of essentialism in the literature is *internalism*. When I initially envisaged my doctoral research project, my impression was that the bogeyman to be exorcized was internalism. I turned to the literature on externalism to arm myself against internalism. However, as I progressed in my research, I came to the conclusion that some of the main externalist views on the boundaries of cognitive systems, while more conducive to my goal to develop a multiscale model of cognition than internalist models, were guilty of proposing a priori boundaries as well. It became clear that any kind of essentialism about the boundaries of cognition is ill suited to the development of an explanatory model able to address multiscale phenomena. Many of my favourite externalists were also essentialists; with a few notable exceptions (Clark, 2008, 2017), they simply privileged different boundaries and levels of description over others. The Equal Partner Principle of radical enactivism (Hutto & Myin, 2013) is perhaps the clearest and most extreme articulation of this view, and claims – in an a priori manner – that the causal factors that lead to cognition and behaviour (in the brain, body, and environment) all count equally and are equally prescient.

This view clearly has significant advantages. However, when read as an *essentialist* claim about the boundaries of all cognitive systems, this seems wrongheaded. Surely, the different causal factors at play in a given process will not (or at least, not always) have the same weight – and not all the factors that might be relevant for cognition in general will turn out to be relevant for every function that is investigated by cognitive scientists. While more amenable to a perspective working across different spatial and temporal scales than a view of the mind as 'what the brain does', then, the radical enactive perspective also makes a priori, essentialist claims about where to draw the relevant boundaries of cognitive systems.

In summary, essentialism in the cognitive sciences is Janus-faced: internalism and externalism are two versions of the same essentialist claim. My contention is that, on either account, we lose much of what contributes to the phenomenon that we call *mind*. To work towards a truly *multiscale* approach to systemic dynamics of cognitive systems, we must

blaze a path *between* the Scylla of internalism and the Charybdis of externalism. To avoid *losing our minds*, losing sight of the full phenomenon of mind and cognition, we must reject any perspective on cognitive systems that privileges one scale or level of description over others, and crucially, we must formulate an alternative explanatory model that can accommodate its full generality.

In place of a metaphysics of emergence, the papers collected in this thesis propose a *formal ontology of multiscale cognitive systems*. As we are using the phrase here, to produce a formal ontology is to employ a formalism – in this case, the mathematical resources made available by the FEP, and more specifically, multiscale constructions of the Markov blanket formalism, which we review in detail below – to answer questions traditionally understood as metaphysical ones; i.e., What does it mean to exists as a system? The advantage of our formal approach to ontology, which we propose as an alternative to the metaphysics of emergence, is that it does not import *a priori* metaphysical assumptions into our study of multiscale systemic dynamics. Instead, it implements a very minimal notion of systemhood, based on the notion of conditional independence, as the main criterion for what counts as a system, and comes as close as we can to letting cognitive systems carve out their own boundaries, via their dynamics.

The four papers collected in this thesis together form a multiscale model of explanation for the cognitive sciences, which serves as the basis for a critique of essentialist models of cognition. This thesis accomplishes its aim by articulating three novel theoretical frameworks to study cognitive systems at and across all scales at which they exist: (1) *variational neuroethology*, a new approach to theoretical neuroethology – the study of adaptive behaviour and the control systems that have evolved to coordinate it – that is grounded in an account of the information-theoretic constraints on cognizing organisms, called the free-energy principle; (2) *variational ecology*, a new take on cognitive ecology that further specifies how hierarchically structured organisms, which craft their own ecological niches, attune themselves to their niche (and vice-versa); and finally, (3) the *cultural affordances framework*, which provides a detailed account of human action, cognition, learning, and culture in terms of environmental affordances and regimes of attention. Based on these novel theoretical pillars, this thesis provides a resolution of the dialectic (and false dichotomy) between internalism and externalism in the philosophy of the cognitive sciences; and provides an account from first principles of where to draw the boundaries of cognitive systems.

5. A comprehensive review of the relevant literature

In this section, I first review broadly the bodies of literature upon which I have drawn in the thesis. When applicable and relevant, I point out the lacunae in each of these bodies, which has motivated the work in this thesis. I then discuss closely related projects in philosophy and the cognitive sciences that have influenced the course of my research and that have converged on similar results.

5.1. Literature review

I have drawn on four main bodies of literature to develop the material contained in this thesis. The first cluster of literature is work on the *variational free-energy principle* (FEP), drawn mainly from computational neuroscience and theoretical neurobiology. The FEP is ultimately the theoretical backbone of this thesis, and does much of the heavy lifting to explain the multiscale integration of system dynamics. The second cluster of literature is work in philosophy and the cognitive sciences carried out from an externalist perspective, that criticizes internalist conceptions of the boundaries of cognitive systems. This literature is subdivided in two groups: work on so-called '*4E approaches*' to cognition – i.e., *embodied*, *extended*, *embedded*, *enactive* approaches to cognitive systems – and the literature on *ecological psychology*. I then discuss some of the literature in cognitive and evolutionary anthropology that motivated the work on human cultural cognition and action that is proposed in this thesis. Finally, the last body of literature upon which I draw is the intersection of the first two: the ecological-enactive approach and what has become known as Bayesian enactivism, or enactivism 2.0.

5.1.1. The variational free-energy principle

The first body of literature upon which this thesis draws is work on one of the fastest growing (and in recent years, most influential) approaches to the study of cognitive systems in the cognitive sciences: the FEP and its corollary, active inference. It is hard to overstate the significance of the FEP for my research agenda in this thesis. Indeed, the FEP is the theoretical lead bearer of this thesis, which enables the formulation of a rigorous multiscale framework to study cognition and adaptive behaviour. The FEP provides mathematical framework and computational modelling tools that, when properly construed, preclude any essentialist interpretations of the boundaries of cognition.

In a nutshell, the FEP is a description of the conditions that must be met by any self-organizing system that exists at nonequilibrium steady state; that is, the necessary constraints

on any system that *exists*, i.e., that maintains itself in a bounded set of states over time and that persists as such a system in the face of environmental perturbations (Friston, 2010, 2013). Importantly, this means that the FEP is formulated for systems that exist far from thermodynamic equilibrium, in what is known as the regime of nonequilibrium steady states (Ao, Chen, & Shi, 2013). Thermodynamic equilibrium is the state of affairs in which the quantity entropy is maximised; where entropy is essentially a measure of the spread, dispersion, or number of configurations in which a system can exist; to be alive entails resisting this entropic decay (Bruineberg & Rietveld, 2014; Jarzynski, 1997; Martyushev & Seleznev, 2006; Tomé, 2006).

To take a simple example, a car engine functions because the degrees of freedom of each of its parts is limited – notably, by the engine design, which limits in the literal sense, in virtue of its physical structure, the possible forms of motion for its parts (e.g., pistons) (Turvey, 1990). Entropy is low in a functioning engine; but higher in a broken engine – one, say, with wobbly pistons. Most systems in nature, from lightning bolts to glasses of sweet soda, inevitably *increase* entropy, and reach equilibrium with their environment when entropy is at its highest. Glasses of cool, fizzy drinks left on the kitchen counter go flat and become room temperature – which is the point at which they reach thermodynamic equilibrium with the surrounding environment. Lightning bolts dissolve the charge gradients that generated them, and thereby equalise the electric charge between the earth and sky. This is true for almost the entirety of natural systems – as most of physics is formulated for closed systems (that do not exchange energy/entropy with other systems) in equilibrium regimes.

The FEP is a story about the remarkable fact that characterises living systems, which is they *resist entropic decay*. The FEP essentially tells us what *must be true of any system with an attracting set*, i.e., what must be true of any system with a *phenotype*, by virtue of that system having a phenotype at all (Friston, 2012, 2013). At its core, the FEP provides an answer to Schrödinger's famous question, What is life? (Schrödinger, 1944).

The FEP provides a crucial information-theoretic measure – called *variational free-energy* – which measures how well a cognitive system is able to fulfil this condition. One crucial aspect of this measure is that it is extensive. In other words, it can be applied in a scale invariant fashion (Dorogovtsev & Mendes, 2002; Mantegna & Stanley, 1995). Scale-invariance here means that the free-energy at each scale can simply be summed or integrated across scales, yielding one single metric for a system across the scales at which it exists. Because this measure is scale invariant, one can model self-organisation at different scales, for different systems, and sum or integrate that measure across scales. Under the assumption

that phenomena across all scales of self-organisation optimise variational free energy, this measure is what allows one to make sense of the manner in which cognitive systems are able to plan and coordinate situationally appropriate behaviour. Variational free-energy, then, provides the key metric to explain how systems can be composed of differentiated systems at each scale, while also evincing *integrated dynamics across scales*; e.g., how an encultured organism – composed of organs, which are in turn composed of cells, while being situated in a given cultural niche and being part of local cultural groups – can *adaptively act* as a *coherent*, *integrated whole*.

One of the central contributions of this thesis is to highlight and discuss the theoretical implications of the scale invariance of variational free-energy for the study of adaptive behaviour in multiscale systems. Notably, this scale-invariance allows one to compare different configurations of different multiscale systems (Friston, Levin, Sengupta, & Pezzulo, 2015; Sengupta, Tozzi, Cooray, Douglas, & Friston, 2016). Without the scale-invariance of variational free-energy, it would be impossible to (1) compare self-organised behaviour at different scales within the same system – e.g., the different contributions of neural behaviour, functional connectivity, organism-environment interaction, and interaction among multiple agents to the overall adaptiveness and appropriateness of behaviour; and it would be impossible (2) to compare different such systems (e.g., two distinct groups or species). This is because without such a measure, there is no 'common currency' or metric between descriptive levels that can effectively track behavioural success at those levels, thanks to which one might make such comparisons. By the same token, without such measure, it is impossible to propose a multiscale model of cognition amenable to empirical testing and which resists essentialist claims about the boundaries of cognitive systems.

The FEP itself says that any system that persists as a system, maintaining itself in a bounded set of phenotypic states, will look as if it is minimising the '*surprisal*' associated with being in a given sensory state (Friston, 2010). Surprisal (aka surprise) in the context is an information-theoretic measure that reflects the probability of being in a surprising state; technically, a state associated with low probability – e.g., a fish out of water. (This informational surprise is not to be confused with the folk psychological notion of surprise.) Variational free-energy is constructed as an upper bound on surprise, based on the (sub-personal, probabilistic) 'beliefs' of an organism about the causes of its sensory states. These beliefs are harnessed in a statistical model called a *generative model* (Friston, Parr, & de Vries, 2017). A generative model is a statistical mapping from causes to observable consequences; it is a graphical model of the causal factors that generate an outcome of

interest (Friston, Parr, et al., 2017). When the variational approach is applied to model cognition and living systems, the relevant generative model is a model of those causal factors that generate sensory outcomes; including, crucially, the actions of the organism itself. It is technically the joint probability distribution that describes the co-occurrence of a sensory state and an external state in the world. In all four chapters, we articulate a statistical conception of the organism's phenotype. We can interpret the generative model as a statistical description of the phenotype, in the sense that it provides a description of the viable couplings between what the organism senses and what causes those sensations. Under the FEP, the adaptive actions of an organism are said to entail a generative model, in the sense that adaptive actions are actions within that mapping between sensation and environment.

The variational free-energy can be read as (negative) evidence for such a generative model; with low free-energy indicating a more highly probable model. This lends the active inference framework the look and feel of Bayesian inference. Systems that obey the FEP via adaptive action are said to engage in *active inference*; since it will look as if such systems are inferring the causes of their sensory states, via the selection of adaptive action policies, i.e., those associated with the least free-energy (Friston, Mattout, & Kilner, 2011; Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2016). Variational free-energy, then, just is a proxy used by organisms for estimating surprise (since it is mathematically constructed as an upper bound on surprisal) – and by tracking this proxy and keeping it in check, cognitive systems are effectively able to track and limit their exposure to the quantity that this proxy tracks – i.e., surprise.

The FEP and active inference are premised on the technical developments and resources of *variational inference*, which is a technique for complex statistical inference, and the *Markov blanket formalism*, which is a formalism that we use to answer the question: What counts as a system? The FEP casts adaptive action as a form of *variational* (approximate Bayesian) inference. The variational methods were originally developed by the physicist Richard Feynman is the context of statistical mechanics (Feynman, 1972). Feynman developed these methods to get around the problems posed by computationally intractable terms that appeared when trying to analytically solve some of the integrals that appeared in the statistical mechanics problems he was addressing. Variational inference involves guessing the parameters of some target probability density function. To do so, we use a *variational density* – i.e., a probability density or (sub-personal) Bayesian belief – which is essentially a guess about the exact form of the distribution we want to approximate. We then use variational inference to improve or tune our guess.

(The exact mathematical details of variational inference are not the focus of this thesis. The best point of entry into the technical literature on variational inference is probably the inaugural work by Feynman (1972). Some tutorials exist that explain applications of the variational formalism in the active inference formalism (Buckley, Kim, McGregor, & Seth, 2017). A monograph is currently being prepared to comprehensively present the technical developments underwriting the FEP (Friston, in preparation).)

The *Markov blanket formalism* is central to the FEP in its use for the study of living, self-organised systems of systems – and it is a central part of the work in this thesis. This formalism was originally proposed by Pearl (1988) in the context of statistical inference. The formalism was later taken up, linked to approximate Bayesian (variational) inference, and applied to model cognitive systems (Friston, 2012, 2013). In a nutshell, the Markov blanket formalism allows one to answer the question – What counts as a system? It thereby provides the basis for a *formal ontology* of living systems, i.e., a theory that answers questions about the *existence* of cognitive systems.

The Markov blanket operationalizes the notion of *conditional independence*. It gives us formal tools to articulate the intuitive notion that, for a system to exist at all as a system, it must be endowed with some degree of conditional independence from its embedding environment. If this minimal requirement is not met, either the set of things we are considering merely dissipates into the environment in which it is embedded – in which case it ceases to be the system that it is – or again, the set is not a system in this robust sense, only a logical sense (we would not, for instance, speak of a set formed by the climate in Montreal, a tree in Central Park, and an ant in Shanghai as a system). Indeed, we (almost never) think of these objects as systems because we have something like 'intuitive feel' for what should count as a thing (Pearl & Mackenzie, 2018). The Markov Blanket formalism is of great help to formalise this intuition – especially how it applies to less obvious cases, such as the ones that are the focus of this thesis (the mind, multiscale cognitive systems, etc.).

A Markov blanket is a set of states that separate or shield some system of interest from its embedding environment. The systemic states thus insulated are known as internal states; and the ones from which they are shielded are referred to as external states. The blanket itself is made up of sensory and active states, which are also defined in terms of their statistical relations; active states are influenced by (but do not directly influence) internal states, whereas sensory states are influenced by external states (but, reciprocally, do not directly influence them). These relations are captured in Figure 1.

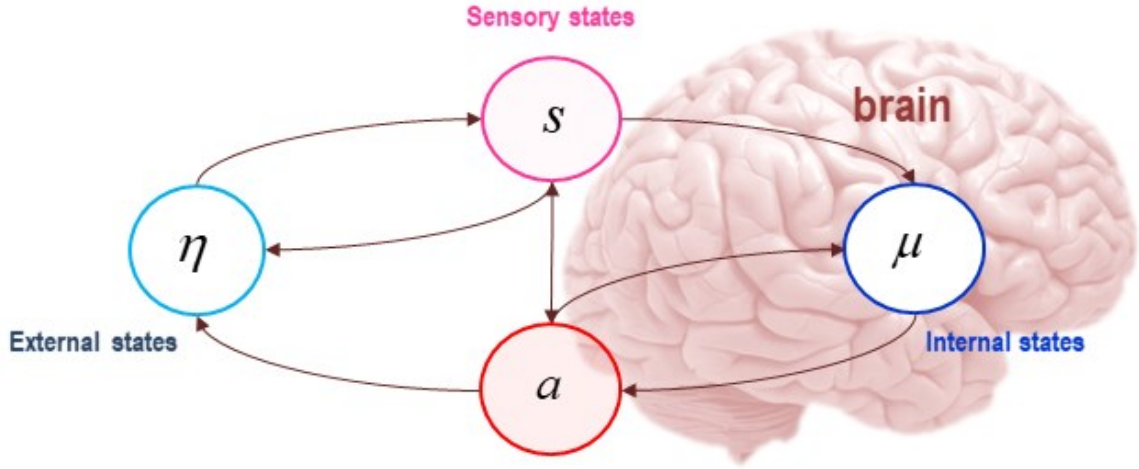


Fig. 1. *A Markov blanket*. Schematic of the quantities that define a Markov blanket and their relations. Here, the Markov blanket is depicted, as well as the system that it enshrouds (the internal states) and the external, non-systemic states. Here, the systemic or internal states, which make up the system in question, are denoted μ ; and the external or hidden states are denoted η . The Markov blanket per se is made up of sensory states, which are denoted by s , and active states, denoted by a . From chapter 4; Ramstead, Kirchhoff, et al. (2019).

In effect, the Markov blanket establishes the *conditional independence* that makes a system count as a system, i.e., as an integrated whole with well-defined boundaries. This separation between internal and external states that is induced by a Markov blanket is meant to be read in a statistical sense; in terms of the propagation of effects or influence from one set of states to another. The Markov blanket 'insulates' internal states from external ones (and vice versa) – in the sense that external states are only able to influence internal states in an indirect manner, namely, via their action on sensory states; and reciprocally, internal states can only affect external ones indirectly, via their effect on active states. Via active inference, over time and on average, it will look as if organisms and their ecological niche were inferring each other's states and statistical properties. On the side of the organism, this will have the look and feel of action, perception, and learning.

The FEP was originally formulated by my mentor and doctoral supervisor Karl Friston as a theory of neuronal processes in the mid-2000s (Friston, 2010; Friston, 2005; Friston, Kilner, & Harrison, 2006). It has proved to be an exceptionally powerful, elegant, and generative framework for neuroscience research. This might suggest that the FEP is an internalist theory of the brain (and nothing more); indeed, many authors have read it to be so (Gładziejewski, 2016; Hohwy, 2014, 2016). In the late 2000s, the explanatory scope of the

FEP was progressively extended, albeit within the confines of computational neuroscience – from a theory of cortical processes (Friston, 2005) to its articulation as a full account of the brain across its different hierarchical levels (Feldman & Friston, 2010; Friston et al., 2006; Kiebel, Daunizeau, & Friston, 2008), to a theory of how adaptive action might be controlled by the brain (Friston, 2010; Friston, Daunizeau, Kilner, & Kiebel, 2010; Friston et al., 2011). This integration of *action* into the mix was crucial, since it showed that the FEP could be more than merely a theory of perception and cognition.

More was to come. In a series of landmark papers from 2012 to 2015, the FEP was extended from a theory of brain function to a theory of living systems in general. In 2012-2013, connections between the FEP and the mathematical apparatus of Markov blankets and Markov decision processes were drawn (Friston, 2012, 2013). This suggested that the FEP could be cast not only as a theory of self-organisation in brains, but more generally as a theory of the class of living system (i.e., a theory of life as we can currently understand it). By 2015, free-energy minimisation was shown to be sufficient to provide an explanation of morphogenesis (that is, cell differentiation). Essentially, via mathematical work and in silico proof of principles, it was shown that – with some very simple assumptions in play, the discussion of which goes beyond what can be discussed in a brief review of the literature – biological self-organisation is an inevitable consequence of the physics and statistical structure of system at nonequilibrium steady state. This, in turn, suggested that the FEP was sufficient to describe the generic necessary constraints on self-organizing and self-maintaining systems (Friston, Levin, et al., 2015). Since the modelling work showed that a higher-order pattern (the target morphology) could be generated as long as free-energy minimisation guided individual components of the system, it was becoming clear that the FEP could model *systems composed of other systems*.

Shortly thereafter, work extending the physics of the FEP showed that the scale invariance of the variational free-energy meant that, at least in principle, one could explain the dynamics of all the nested systems and their integration across scales (Sengupta et al., 2016). It was hinted that this could lead to a new take on the study of adaptive behaviour and its control (neuroethology) – although the details of this proposal would need to be worked out. This is where the work in the present thesis comes into play. As we examine in the next main section of this thesis, one of the original contributions of the present work is to systematically articulate the project of a variational neuroethology, and to further extend the FEP to phenomena beyond the brain.

Another aspect of the FEP that had yet to be exploited was its *symmetry*. Active inference tells us both how organisms will adapt to their niches, and how ecological niches will slowly change to match the structure of their denizens. Mathematically, the FEP is also a story about how *organism-niche systems* attune themselves to one another; in the sense of being statistically informative or predictive of each other, and entailing the same predictive consequences. This implicit symmetry is developed in the third chapter of this thesis.

5.1.2. 4E cognition and ecological psychology

This thesis draws from an extensive and growing externalist literature, which has conducted a systematic critique of internalism in quite some detail over the last two decades. The theoretical and philosophical work in this literature was crucial to helping extend the FEP into a theory of adaptive multiscale dynamics and cognitive ecology. These approaches, conducted in the spirit of externalism, fall under the rubric of '4E cognition'. The 4 Es in question are the 'embodied', 'extended', 'embedded', and 'enactive' conceptions of the mind; for an overview, see (Kiverstein & Clark, 2009; Newen et al., 2018). Another E that should be considered here, and which informs much work in this thesis, stands for 'ecological psychology' (Chemero, 2009; Gibson, 1979).

5.1.2.1. The embodied mind

The embodied mind is the view that cognitive processes depend on the workings of the body; and that information processing is often 'outsourced' or 'offloaded' to (or, more accurately, directly processed by) the morphological and anatomical structure of the body. Although there is much heterogeneity in this literature, the fundamental claim is that cognitive processes depend on the processes and physical structure of the body for aspects of their underlying structure and function. The foundational work in this area is work by philosophers and cognitive scientists Gallagher (2006); Noë (2004); Varela et al. (1991).

The least controversial way that this view on cognition has been pursued is the suggestion that the dynamic, adaptive actions of cognitive systems can leverage the physical structure and processes of the living body to perform cognitive work. Some have cast this as 'offloading' cognition to the morphological structure of the body; but, as my mentor and Master's supervisor Pierre Poirier has pointed out to me in discussions, it is more accurate to say that those aspects of cognitive processing that were already done by the body do not have to be uploaded to the brain. Indeed, as the brain is more plastic than body, it would be surprising if its processes were the ones being offloaded to anything. The structure of the

body itself, in such cases, limits the degrees of freedom required to control and guide movement, which allows organisms to save on the metabolic costs of neuronal information processing (Turvey, 1990). This is known as morphological computation (Pfeifer & Gómez, 2009; Pfeifer, Iida, & Gómez, 2006), and inspired a new line of research in robotics, premised on the idea that internal representation was not necessary for – at least minimal forms of – adaptive behaviour (Brooks, 1991; Chiel & Beer, 1997; Paul, 2006).

The embodied mind also addresses the embodied character of lived experience, drawing on phenomenological philosophy (Husserl, 1990; Merleau-Ponty, 2013/1945). Work on embodied metaphors suggests that sensorimotor processes – and the basic layout and disposition of our bodies when engaged in intentional action – provide a structure that scaffolds, and provides the basis from which can develop, more abstract forms of thought and experience (Kirmayer & Ramstead, 2017; Lakoff & Johnson, 2008). Work on embodied cognition has also addressed the ways in which the body is a basis and medium for rich social and cultural communication and self-expression (Froese & Di Paolo, 2011).

Historically, the embodiment paradigm was the first E-approach to emerge. The embodied mind was a controversial position at the time it was proposed. Indeed, the dominant approach to the study of cognition in the late 1980s and 1990s was symbolic computationalist cognitivism; namely, the Fodorian view that cognition was mental computation, i.e., the rule-based manipulation of mental symbol-like representations (Fodor, 1975; Fodor, 1983). Varela and colleagues were the first to propose a systematic framework to challenge the dominant paradigms of the time (Varela et al., 1991). Today, the embodied approach is largely standard textbook material. Indeed, work in robotics has demonstrated the utility of the approach (Beer, 1995; Chiel & Beer, 1997).

The relevance of the embodied approach to cognition for the thesis is multifaceted. In extending the FEP to phenomena beyond the brain, this thesis essentially espouses the view that it is the entire phenotype that engages in cognitive activity. In other words, all the processes of the body, whether homeostatic or allostatic, participate in the generation of adaptive behaviour.

The first and second chapters of this thesis show how to formulate mathematically the claim that the body is involved in the same game as the brain, just at another set of spatial and temporal scales. We are essentially providing an information-theoretic framework for the embodied mind.

5.1.2.2. The enactive approach

The embodied mind paradigm dovetails nicely with the enactive approach to living systems (Di Paolo & Thompson, 2014; Thompson, 2010; Varela et al., 1991). The enactive approach covers many different (and sometimes even opposed) approaches to cognition. The central claim of all brands of enactivism is the pragmatist emphasis on adaptive action: meaning is processual (Gallagher, 2006, 2017; Hutto & Myin, 2017; Hutto & Myin, 2013; Noë, 2004; Thompson, 2010; Varela et al., 1991). In other words, for an organism to make sense of itself and its world is a process, and meaning is brought forth into the world through patterned loops of adaptive, purposeful action.

The enactive approach coheres broadly with the embodied mind, in that all enactive approaches also endorse the view that cognition is embodied – however, there are approaches to embodied cognition that reject the radicalism of enactivism. The enactive approach is still hotly debated and controversial – and much of this thesis is based on the implications of taking it seriously, while pushing back against what transpires today as some of its more conservative aspects.

According to the useful typology proposed by philosophers Hutto and Myin (2013), there are three main forms of enactivism, to which I and others have added – in part through work contained in this thesis – a fourth form. These are: autopoietic, sensorimotor, radical, and our novel Bayesian approaches to enactivism. *Autopoietic* enactivism is a form of enactivism that tries to ground the cognitive activities of living organisms in the self-production (aka autopoiesis) of living organisms (Thompson, 2010; Varela et al., 1991). . Autopoietic enactivism proposes a theory of the kinds of general system-level features that a network of processes must exhibit to be self-creating in this fashion (Di Paolo, Buhrmann, & Barandiaran, 2017; Di Paolo & Thompson, 2014; see Figure 2). This is also the original brand of enactivism, proposed contemporaneously to the embodied mind.

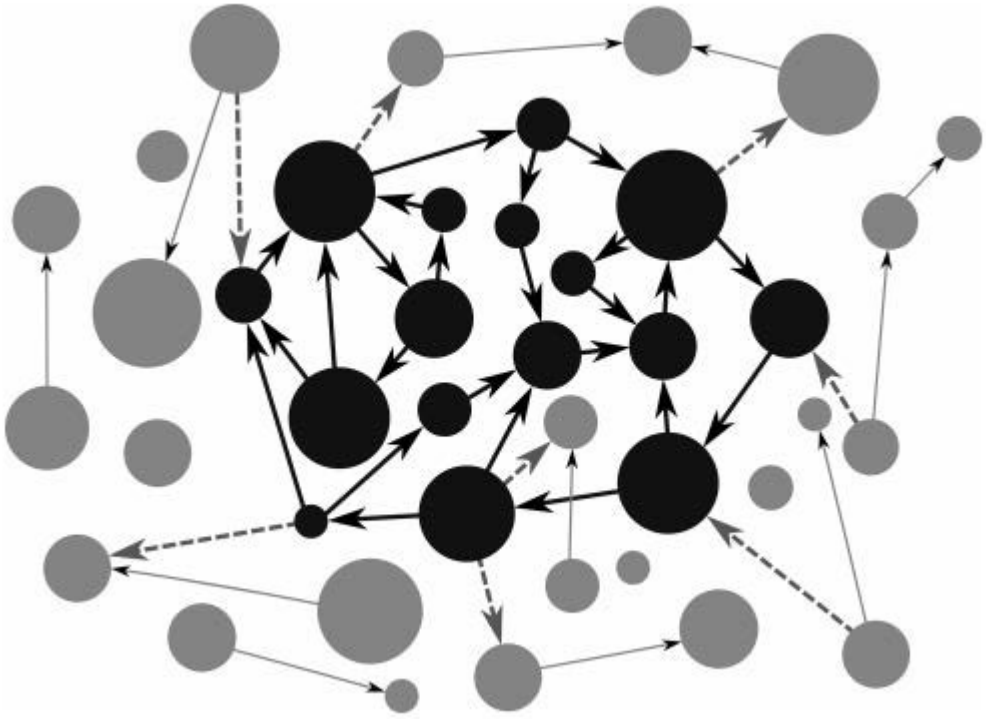


Fig. 2. *Operational closure and autopoiesis*. Autopoietic enactivism foregrounds the notions of autopoiesis and operational closure. Autopoiesis is defined as the capacity of a system to autonomously generate and maintain its component parts and processes, in the face of environmental perturbations. Operational closure is defined as the organisational form of the processes that enable autopoietic self-production and the conservation of the systemic boundaries via interdependent processes. Modified from Di Paolo and Thompson (2014).

Sensorimotor enactivism emphasizes the fact that cognitive processing exploits the tight coupling between the statistical profile of actions and their sensory consequences – what have been dubbed sensorimotor contingencies (Noë, 2004). *Radical* enactivism pursues the claim that 'the world is its best representation'; i.e., that basic forms of cognition do not involve internal representations (Hutto & Myin, 2017; Hutto & Myin, 2013). That is, organisms need not reconstruct an internal representation of an external world to guide action, but rather can act on the basis of sensorimotor contingencies to guide adaptive action. I examine the fourth flavour, Bayesian enactivism, below.

Despite recognising that cognitive systems extend beyond the confines of the brain, some of these views remain essentialist; they define the workings of the mind as essentially being this or that kind of thing – only now, the locus of explanation has shifted away from the brain, towards other factors. The problem with many proposals in the autopoietic, sensorimotor, and (especially) radically enactive approaches is that they end up rejecting the

most essential, generative, and productive tools that are available to cognitive scientists, namely, *computational methods* and *information theory*. This trend in the enactivist literature began with work on minimal forms of cognition, in which minimal systems were designed that evinced adaptive behaviour but did not require explicit internal representations of an external world – and literature that emphasised the dynamic nature of cognition, rejecting the metaphor of the brain as a digital computer (Brooks, 1991; Thelen & Smith, 1996; Van Gelder, 1998). This criticism has been raised by other 4E theorists, who espouse embodied, extended, or embedded cognition, but reject the radical anti-internalism of some enactive theorists (Chemero, 2009; Clark, 2015a).

Enactivism was a very progressive position in the late 1980s and 1990s. However, many of the questions that concerned enactive theorists (namely, "Is cognition a form of computation?") have been rendered moot by the development of computational neuroscience. Computational neuroscience is the use of computational methods and information theory to model the dynamics of cognitive systems. Whereas the first generation of enactivist theory, formulated in the 1990s and early 2000s, was concerned with showing that cognition was not merely symbolic or digital computation, computational neuroscience employs computational methods as tools to investigate the behaviour of cognitive systems – and does not (or at least, not automatically) come pre-packaged with metaphysical assumptions about the computational nature of cognition. Admittedly, some of the writing in this tradition speaks of 'neural computation' (Piccinini, 2015); but often the use is deflationary, and stands as a shorthand to express the way patterns of information are structured and restructured by cognitive processes and action (Bruineberg, Kiverstein, & Rietveld, 2016; Kirchhoff & Robertson, 2018).

5.1.2.3. Extended and embedded (and extensive) cognition

Extended cognition and embedded cognition are two versions of roughly the same claim, namely, that elements external to the boundary of the skull or skin can participate in cognition. Extended cognition is the claim that cognitive processes extend literally out into the environment, such that environmental parameters can be counted as part of the realization base of cognitive processes (Clark, 2008). On this view, the boundaries of cognitive systems are not constrained by the boundary of the skull.

Embedded cognition is the (much less radical) claim that cognitive processes do not literally extend out into the world, such that the boundaries of cognitive systems would extend to elements in the environment, but rather that cognitive processes merely *recruit*

components that extend their innate, intrinsic information processing capacities and provide them with abilities they wouldn't otherwise have (Rupert, 2004, 2009). For the purposes of this thesis, we will not delve into this distinction, and will speak of extended cognition to cover both families of theories. The foundational papers on extended cognition is work by philosophers Clark (2008) and Clark and Chalmers (1998); see Adams and Aizawa (2001); Adams and Aizawa (2009); Adams and Aizawa (2010) for a critique.

Extensive cognition is the newest of the E approaches that I review. It was proposed by philosophers Dan Hutto, Michael Kirchhoff, Erik Myin, and Julian Kiverstein (Hutto, Kirchhoff, & Myin, 2014; for a closely related discussion, see Kirchhoff & Kiverstein, 2019). Work in this thesis builds on the notion of extensive cognition. This is the view that the boundaries of cognitive systems are not fixed once and for all, but rather are negotiated dynamically and are open to change. On this view, the extended and embedded mind start from a false assumption; namely, that cognition is something that initially is confined within the organism or the brain, and that needs to be extended into the world. Proponents of extensive cognition instead argue that cognition starts out as an enactive or pragmatic engagement with the world. It is only in fairly recently evolved creatures like ourselves, and in animals capable of 'offline' cognition (i.e., thinking about states of affairs that do not currently obtain) that thinking becomes internalised as such (Hutto et al., 2014; Kirchhoff & Kiverstein, 2019). On this view, which is in some ways continuous with (but less extreme than) radical enactivism, it is only through the process of *enculturation* that cognitive processes acquire content (that is, becomes subject to truth or satisfaction conditions) – and thereby can be said to represent states of the world (Hutto et al., 2014).

While the extensive approach to cognition is important to the theoretical framework that is developed in the fourth chapter, another approach to the way cognition involves the world arguably had a greater influence on my work – namely, ecological psychology.

5.1.2.4. Ecological psychology

Ecological psychology theory is continuous with many of the desiderata of the enactive approach, but predates them by quite some time. Ecological psychology is a research program in psychology that begins with a critique of the starting point of much of the cognitive sciences; namely, the assumption of poverty of stimulus. It relies on the construct of affordances, proposed by Gibson (1979) – which can be cast as a corrective to the assumption of the poverty of stimulus (Anderson, 2017). Chomsky (1965) conjectured that infants were not exposed to a rich enough stimulus to learn language. This 'poverty of

stimulus' claim was used to argue that, since the stimulus was poor in information, stimuli needed to be augmented via the resources of an internal mental model or representation.

Contrariwise, Gibson argued that the stimuli to which the organism is exposed are extremely rich, when we consider their *dynamics* – that is, the way they unfold over time. On this view, the isolated, point-like stimulus or sense datum is an unrealistic idealization; in reality, sensory surfaces are dynamically exposed to the environment. The relevant patterns of stimulation that are directly readable from the sensory array of organisms carry information that can be leveraged to guide adaptive behaviour; these patterns are called affordances in classical, Gibsonian ecological psychology.

In recent years, Gibson's original sensory-perceptual concept has been transformed and broadened – perhaps to the chagrin of some ecological psychologists (Heras-Escribano, 2019a, 2019b), who disagree that such a conceptual shift is consistent with Gibson's original formulation of affordances. Although the new construct admittedly bears little resemblance to Gibson's original one, its usefulness for our purposes is significant. Under the impetus of work from cognitive scientists and philosophers – such as Chemero (2009), Bruineberg and Rietveld (2014), and Rietveld and Kiverstein (2014) – affordances have been redefined more broadly as possibilities for action in the environment. Chemero (2009) was the first to propose to ground enactive theory on the well-established results of ecological psychology. Affordances, on this view, are interactional properties between the abilities had by agents and the relevant features of their shared material environment.

This thesis draws on a recent synthesis of ecological psychology with the enactive approach and the free-energy principle, called the skilled intentionality framework (SIF) (Bruineberg & Rietveld, 2014; van Dijk & Rietveld, 2016). Building on previous work bridging ecological and enactive perspectives Chemero (2009), the SIF uses the resources of enactivism and ecological psychology to suggest that organisms engage with a field of affordances, specified in terms of salient environment features and the abilities with which organisms are endowed. The perspectives of ecological psychology, and the SIF in particular, loom large in this thesis.

There are two main blind spots in this literature, which this thesis attempts to address. The first, which arises when the SIF is applied to species such as humans, which exist in complex social and cultural niches, is the question of how best to approach affordances that are particular to a given culture. Promising work had been accomplished in this direction (Rietveld & Kiverstein, 2014), but was not yet brought into contact with new work in cognitive and evolutionary anthropology on culture-gene coevolution (Henrich, 2015) and the

cooperative nature of human phenotypes (Hrdy, 2011; Tomasello, 2014). The first chapter of this thesis extends the SIF via this body of work, to provide an account of specifically *cultural* affordances and their mechanistic basis.

Another central limitation of ecological approaches, as originally formulated by Gibson (1979), is the lack of a mechanistic theory to explain how organisms engage with affordances. Follow up work (Chemero, 2009) remediates this by integrating ecological psychology with embodied and enactive conception of mind. Newer work has suggested that Bayesian approaches to cognition might be apt to ground affordances mechanistically (Clark, 2015b). This thesis further remediates this lack of a mechanistic basis by providing a variational physics of affordances, in the second and third chapters.

5.1.3. Culture-gene coevolution, shared attention, and the cooperative turn in cognitive and evolutionary anthropology

Work on human cultural dynamics in this thesis, and especially in the first chapter, rests on new developments in cognitive and evolutionary anthropology that foreground (1) the fact that human culture and human genetics have coevolved and (2) the manners in which collective allocations of attention and salience pattern the ways in which human agents are enculturated. Indeed, the first chapter is an attempt to integrate these novel findings and frameworks with the bodies of literature just reviewed, in order to propose a novel account of human cultural action and cognition informed by the most recent developments in cognitive and evolutionary anthropology, broadly construed. When I met him in 2015, my mentor, close friend, and collaborator Professor Samuel Veissière had been working on the ways in which patterned allocations of attention that were relevant to understanding human culture, and on the question of culture-gene coevolution. Professor Veissière's work informs the first chapter of this thesis greatly.

Concerning the first part, this thesis mainly draws from the work of evolutionary biologists and evolutionary anthropologists such as Joseph Henrich (Henrich, 2015; Moya & Henrich, 2016) and Sarah Hrdy (Hrdy, 2011), who examine the manner in which human cultures emerged from deep historical processes rooted in human biology. This thesis as a whole also draws heavily the work of my mentor and doctoral supervisor, psychiatrist Laurence Kirmayer (Kirmayer, 2008; Kirmayer, 2015; Kirmayer & Bhugra, 2009; Kirmayer & Gold, 2012). These theorists claim that human biology is fundamentally shaped by culture; and that, reciprocally, human culture cannot be understood without understanding its anchoring in human biology. The work of Professor Kirmayer also draws on embodied and

enactive approaches to the cognitive sciences, making this approach even more amenable to the aims of my thesis.

The framework that is leveraged in the first chapter for thinking about the effects of shared attention on sociality and culture find their origin are so-called 'relevance models' of cultural cognition. These models, building on the work of Grice (1989), emphasise that human cognition mostly proceeds on a 'scan and stop' basis: human agents secure cues that indicate high reliability information, and stop thinking once such cues are secured (Wilson & Sperber, 2002, 2012). These models are broadly consonant and complementary to other, more theoretically-oriented work in anthropology, focusing on the relations between the public order and the ways that enculturation fashions our patterns of attention; namely, the theories of civic inattention of Erving Goffman (1971) and the construct of habitus from the work of Pierre Bourdieu (1977, 1984). More recent work on this comes from psychologist Michael Tomasello (2014), who emphasises the role of the joint and shared forms of attention in human development. For instance, Tomasello notes that the education of human infants almost invariably involves situations where parental figures share a common attentional object with their child, which plays a critical role in their development.

This thesis mainly uses this body of literature to correct conceptions of human culture in other bodies of literature that are drawn upon – rather than criticize its lacunae directly. In particular, the embodied-enactive conception of mind tends to be very thin in its treatment of social and cultural phenomena – with a few notable exceptions (Fabry, 2017; Gallagher & Ransom, 2016; Rietveld & Kiverstein, 2014). Even when they do examine culture in detail, however, the new developments just reviewed are seldom taken into account. We correct this situation in the model that we propose in the first chapter, which guides subsequent investigations.

5.1.4. Bayesian enactivism and enactive inference: The enactive approach meets active inference

There has been a growing *intersection* of the two first bodies of literature just reviewed; more specifically, at the intersection of the FEP and embodied, enactive, and ecological approaches to the dynamics of cognitive systems. The first paper to propose such a convergence was Bruineberg and Rietveld (2014). They proposed an integrated model combining aspects from ecological psychology, the enactive approach, and the FEP. Karl Friston, along with philosophers and cognitive scientists Michael Kirchhoff, Tom Froese, Shaun Gallagher, Micah Allen, and Michael Anderson followed up these ideas in a series of recent papers on

the relations between the FEP and the commitments of autopoietic variants of enactivism, arguing that the FEP absorbs, explains, or supersedes autopoiesis (Allen & Friston, 2016; Anderson, 2017; Gallagher & Allen, 2016; Kirchhoff, 2016; Kirchhoff & Froese, 2017).

My thesis work exemplifies and extends this approach. Indeed, the papers collected in this thesis propose a novel, Bayesian form of enactivism. Bayesian enactivism is an articulation of enactivism for the age of computational neuroscience. Its relation to the other forms of enactivism is as follows: Bayesian enactivism extends sensorimotor enactivism by showing what kind of minimal cognitive information processing machinery is necessary for the intelligent engagement with sensorimotor contingencies (Anderson, 2017; Engel et al., 2016). Bayesian enactivism absorbs or supersedes autopoietic enactivism in that it provides an explanation from first principles of the construct that was its basis; namely, we provide an explanation for the autopoiesis of living systems via active inference and Markov blankets. Bayesian enactivism rejects radical enactivism as, perhaps ironically, being too conservative; it foregoes the most powerful tools available in the cognitive sciences for the sake of ideological purity.

In rejecting the tools that made internalism a powerful paradigm for neuroscience research, radical externalists were depriving themselves of some truly revolutionary new theories and tools. Some of the most interesting and powerful tools to study the mind were originally developed in an internalist explanatory setting; e.g., information theoretic tools for statistical analysis and computational modelling. This thesis aims to advocate methodological pluralism as well, premised on the ontological pluralism that follows from a radically multiscale approach to the study of cognitive systems.

The final chapter of this thesis proposes a novel reading of embodiment and enactment predicated on the FEP. The idea that is proposed rests on the construct of generative models from the variational formulation. We propose an interpretation of enactment as the bringing about of the phenotypic states (i.e., the enactment of a generative model) and of embodiment as the encoding of a variational model, or best guess. We thus integrate enactive, embodied, and ecological views on cognitive systems via a multiscale articulation of the FEP.

5.2. Contemporaries and closely related work

It should be acknowledged that the ideas and original contributions made in this thesis explicitly draw on recent work by philosophers Jelle Bruineberg, Axel Constant, Tom Froese, Michael Kirchhoff, Julian Kiverstein, and Erik Rietveld. Some of the work presented here

converged on similar results with work by these authors. This is not surprising. Almost all of us have, at one point or another, worked at University College London with the Theoretical Neurobiology group directed by Professor Karl Friston, who first proposed the FEP a decade ago. Indeed, in many cases, Professor Friston introduced us to each other.

This work builds on the pioneering work of McGill researchers Laurence Kirmayer and Samuel Veissière. Professor Kirmayer blazed the path for a cultural neuro-phenomenological approach to the study of mind and mental health, which was the main inspiration for this doctoral project. This thesis would not have been possible without his influence – and indeed, without his help. Professor Veissière first introduced the notion of cultural affordances that we use in the first chapter in talks and seminars from 2014-2015, which had a profound effect on me. Professor Veissière also was the first to theorize the role of regimes of attention, as we employ the construct, in his novel work on internet subcultures circa 2015 (Veissière, 2015, 2016).

This thesis is indebted to, and draws heavily on, the work of philosopher Jelle Bruineberg. Bruineberg and his collaborators at the University of Amsterdam, Professor Erik Rietveld and Professor Julian Kiverstein, were, in 2014, the first to propose an integrative framework combining enactive and embodied approaches to cognition, ecological psychology, and the free-energy principle – the SIF, which we reviewed above. This thesis is deeply indebted to their pioneering work. As we state in the second chapter, the variational ecology that is proposed there integrates the SIF with the variational neuroethology presented in the first chapter and the variational approach to niche construction.

Work in this thesis is deeply indebted to my close friend and collaborator, the cognitive scientist and philosopher Axel Constant. We developed the variational approach to niche construction during his MA studies (Constant, Ramstead, Veissière, Campbell, & Friston, 2018), which he undertook under the supervision of Sander van de Cruys and Maren Wehrle at the Catholic University of Leuven. The variational approach to niche construction – that we developed based on pioneering work by Jelle Bruineberg in 2017 and 2018 – was one of the central components of the work that would lead to the development of the variational ecology that is presented in the third chapter of this thesis. The third chapter integrates the variational neuroethology, which I developed in 2017 and 2018 with Paul Badcock and Karl Friston, with the variational approach to niche construction, to provide a fully generalizable framework for modelling living, cognitive systems interacting with their ecological niches across scales – the variational ecology.

Philosophers Michael Kirchhoff and Tom Froese were the first theorists in the enactive approach to systematically explore the relation between the free-energy principle and the Markov blanket formalism and the commitments of the enactive approach. It should be noted that Kirchhoff was involved independently in research on nested Markov blankets, around the same time that I was working through the implications of multiscale modelling of adaptive action and cognition via the nested Markov blankets formalism with Paul Badcock and Karl Friston, circa 2016-2017. Kirchhoff's research visit at University College London allowed him and his colleagues to approach the question of nested Markov blankets from the point of view of generative models and their relation to phenotypes – as opposed to approaching it from the question of spatial and temporal scales, which was our starting point. Our results, and the resulting philosophical perspectives, have converged. We have since become friends and collaborators. Kirchhoff is the second author on the last chapter of this thesis.

6. Contribution to original knowledge

The main original contribution to knowledge made by this thesis, which is taken up over all four chapters that make it up, is twofold: (1) the proposal of novel theoretical frameworks – and, in some chapters, of corresponding mathematical models – to study cognitive systems, especially humans and human societies, at and across all the spatial and temporal scales at which they exist; and (2) the consideration of the implications of these multiscale models for the debate in the philosophy of science over where to draw the boundaries of cognitive systems. The thesis leverages the FEP, embodied-enactive-extended approaches in the cognitive sciences, and ecological psychology to formulate three new theoretical frameworks for multidisciplinary research on multiscale systems in the cognitive sciences, called the *cultural affordances framework*, *variational neuroethology*, and *variational ecology*.

The novel contribution of these frameworks is to enable those working in the cognitive sciences (1) to model, and indeed explain from first principles, the cognitive dynamics of living systems beyond the scale of the brain – from cells the societies; and (2) to provide a theoretical scaffold to enable novel, multidisciplinary research in the cognitive sciences. Having established these theoretical results, the thesis then attempts to draw on them to dissolve the very terms in which the debate is posed in the philosophy of the cognitive sciences between internalism and externalism. The thesis shows that internalist and externalist approaches both err in their essentialism about the boundaries of cognitive systems; and it celebrates a pluralistic approach to the ontology and study of cognitive systems. More specifically, this thesis makes four main original contributions to knowledge:

6.1. A multiscale perspective on the study of human sociocultural action and cognition, via a cultural reading of affordances: The cultural affordances framework

The first chapter, "Cultural affordances: Scaffolding local worlds through shared intentionality and regimes of attention," was published in *Frontiers in Psychology* (Ramstead et al., 2016). It was written in 2015-2016 and constitutes my first attempt to work out a multiscale perspective on nested cognitive systems; in this case, perhaps the most interesting such systems, human sociocultural groups. More specifically, the chapter provides a multiscale perspective on the problems posed by the acquisition of cultural norms, values, and practices.

The chapter synthesises much of the extant literature on human cultural learning and proposes a tractable mechanism to explain human enculturation. The chapter revolves around the construct of *affordances*, which we borrow from ecological psychology. An affordance is

a possibility for engagement with the environment – it is a relational feature that binds an organism to its environment. One of the novel contributions of this chapter is to articulate a framework to theoretically approach the kinds of affordances that populate the human ecological niche, which we call 'cultural affordances.'

The mechanism that we propose in this chapter integrates causal factors that are internal to the organism, notably, brain-based mechanisms of prediction and attentional modulation; as well as factors external to it, namely, the immersive participation of an agent in patterned cultural practices. The mechanism is depicted in Figure 3.

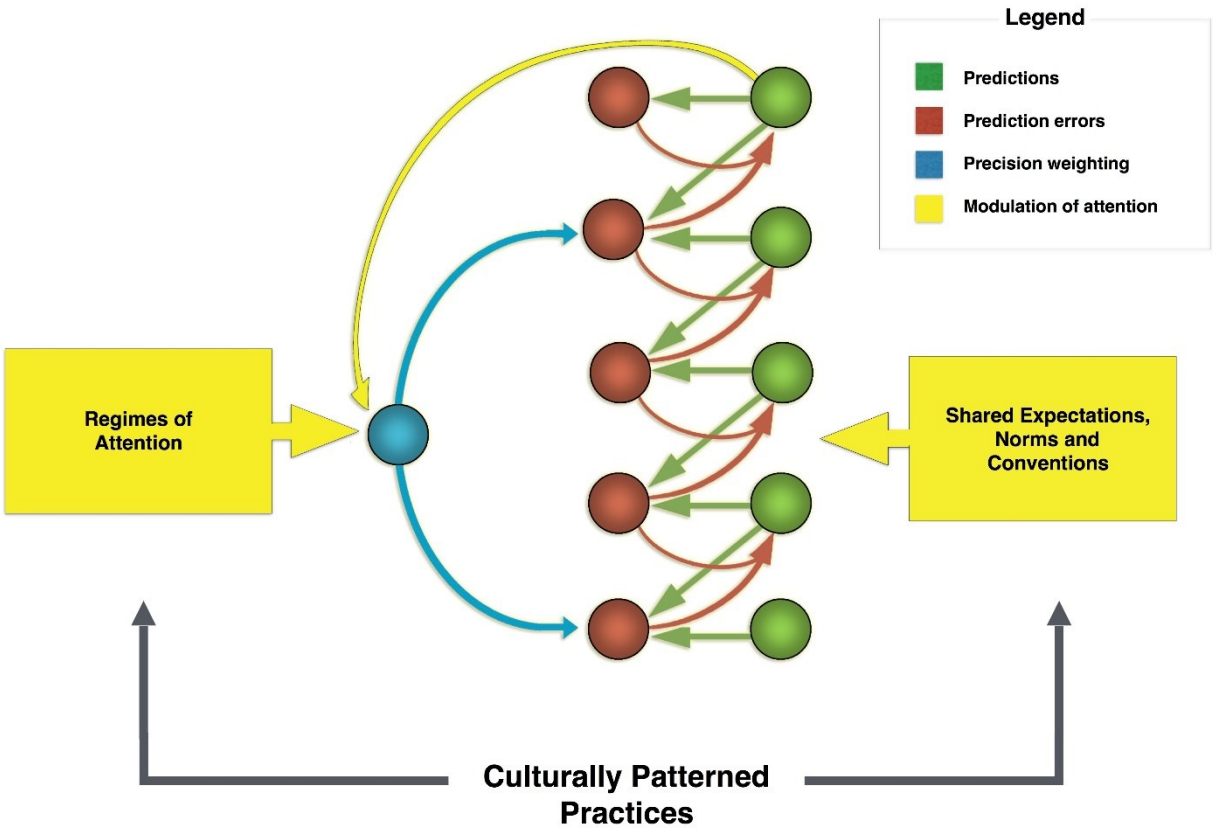


Fig. 3. *A diagram of the mechanism that enables the learning of cultural affordances.* What we call 'regimes of attention' are the patterned cultural practices that guide agents to attend to this rather than that. The model draws on the free-energy principle and in particular in the way that attention is theorized in that model. Under the free-energy principle, attention has the function of gating or modulating the deployment by agents of perception-action loops. By directing attentional processes, regimes of attention end up enculturating agents with the predictive expectations that enable enculturation in a given a sociocultural group. From Ramstead, Veissière, & Kirmayer (2016).

The claim is that humans become attuned to the cultural affordances of their sociocultural niche thanks to the patterning of attention that is made possible by immersive participation in certain kinds of patterned cultural practices, which we call 'regimes of attention'. A regime of attention is a practice that signals the salience (or lack thereof) of certain kinds of stimuli and situations. The interaction between regimes of attention and brain-based mechanisms enable the attunement to cultural affordances. The cultural affordances model thus provides a multiscale framework for the study of human social and cultural cognition.

6.2. An articulation of the FEP beyond the brain: Variational neuroethology and variational ecology

One of the main contributions of this thesis is to extend the FEP – which was developed as a unified theory of the brain – to systems beyond the brain, and to show that the FEP provides the cognitive sciences with the formal mathematical framework required to articulate an integrated, multi-scale perspective from which to study living systems that exist across multiple spatial and temporal scales. This is realised especially in the last three chapters. The two middle chapters (the second and third chapters) provide the technical results required to accomplish this extension, and the third discusses the broader philosophical implications of this extension.

The second chapter, called "Answering Schrödinger's question: A free-energy formulation," is a review paper that was published in the journal *Physics of Life Reviews* (Ramstead et al., 2018a) with 14 peer commentaries (Allen, 2018; Bellomo & Elaiw, 2018; Bitbol & Gallagher, 2018; Bruineberg & Hesp, 2018; Campbell, 2018; Daunizeau, 2018; Kirchhoff, 2018; Kirmayer, 2018; Leydesdorff, 2018; Martyushev, 2018; Pezzulo & Levin, 2018; Tozzi & Peters, 2018; Van de Cruys, 2018; Veissière, 2018). We were invited to write a response (Ramstead et al., 2018b) as well.

"Answering Schrödinger's question: A free-energy formulation" explicitly addresses the question of nested temporal and spatial scales, and is the first paper in the literature to argue that the FEP can systematically be used to articulate a theory of adaptive behaviour that can explain the dynamics of cognitive systems across all the scales at which they exist. It proposes a novel synthesis to study the adaptive behaviour of nested systems of systems. It does this by drawing on some of the most influential models in theoretical biology from evolutionary systems theory (EST) (Depew & Weber, 1995; Fisher, 1930; Kauffman, 1993; Sarkar, 1992; Wright, 1932). EST is a synthetic discipline that studies the intersection of self-organisation and generalised selection (of which natural selection is just one subtype)

processes in the biological sciences. The most crucial of these frameworks for our purposes is Tinbergen's four seminal research questions (Tinbergen, 1963) – i.e., mechanism, ontogeny, phylogeny, and function – which we combine with the FEP.

More specifically, the main contribution to original knowledge of the first chapter is the proposal of a new approach to the study of adaptive behaviour and its control in nested living systems called *variational neuroethology* (VNE). As a framework, VNE has two main components: (1) a multi-scale ontology of nested cognitive systems (i.e., systems of systems); as well as (2) a multidisciplinary research heuristic to study the behaviour of nested cognitive systems. Neuroethology is the study of the evolution and dynamics of animal behaviour and their associated control systems (i.e., their brains). VNE builds on the FEP and on Tinbergen's four seminal research questions in biology to propose a fully generalizable framework to study adaptive behaviour at and across spatial and temporal scales. The original theoretical development of this chapter is to articulate a theoretical model, accompanying mathematical formalism, and multidisciplinary research heuristic to enable cognitive scientists to model nested cognitive systems; that is, to model the dynamics of systems that are themselves made of other systems (e.g., cells that are nested in tissues, themselves nested in organisms, embedded in a shared ecological niche).

The central contribution of the second chapter to the theoretical framework of the FEP is to show that any living system can be described as recursively nested Markov blankets; that is, *Markov blankets of Markov blankets* – all the way up, and all the way down. The crucial observation that underwrites this development is that the components of Markov blankets are themselves systems, which means they also must have a Markov blanket; and the interactions between blankets induce a sparsity structure that leads to blankets of blankets. This recursively nested structure of blankets is depicted in Figure 4, below. The main theoretical contribution of the first paper is to theoretically unpack this nested structure of systems of systems.

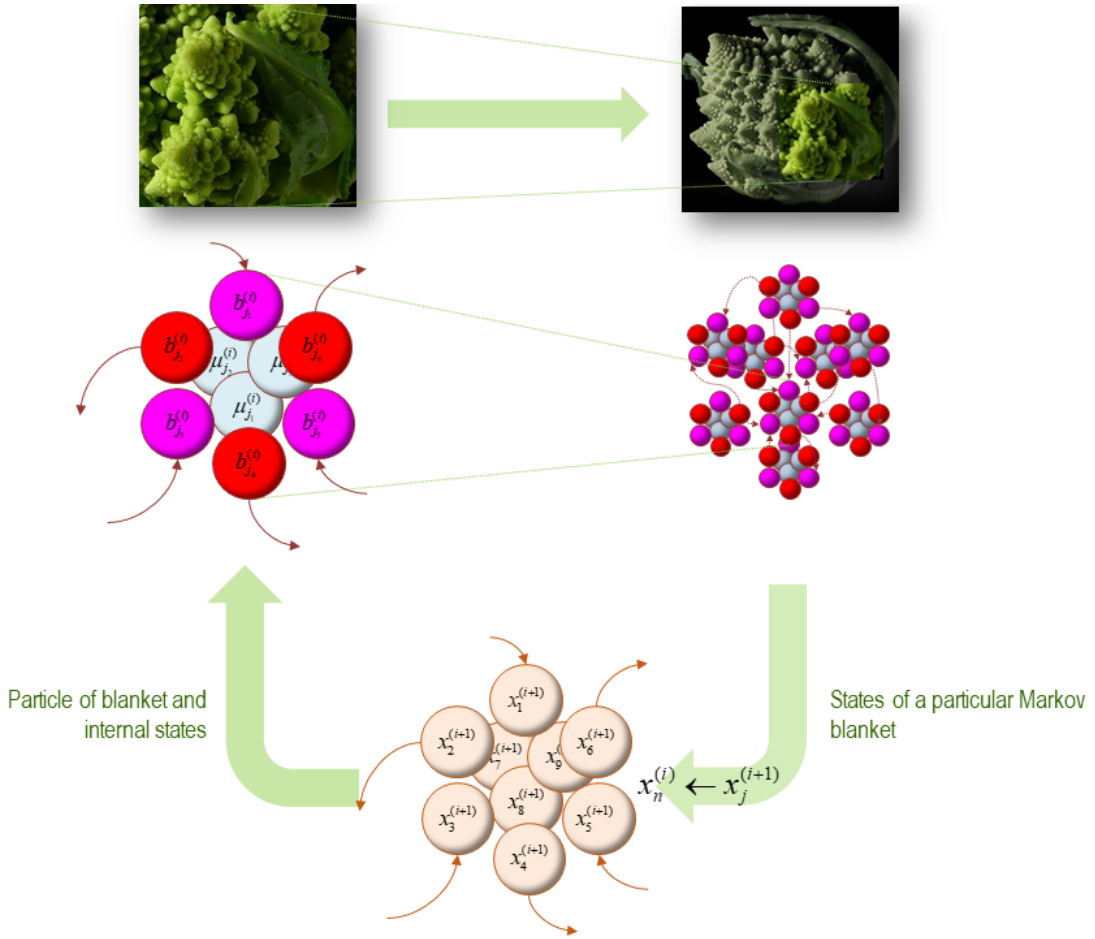


Fig. 4. *Blankets of blankets*. This schematic figure depicts nested Markov blankets of Markov blankets. More specifically, the figure illustrates how to move from one scale of Markov blankets, with its blanket states ($b_j^{(i)}$) and internal states ($\mu_j^{(i)}$), to the next. See the second and third chapters for technical details. From chapter 3; (Ramstead, Constant, et al., 2019)

The third paper in this thesis, "Variational ecology and the physics of sentient systems," is a follow-up paper – a direct sequel to the first chapter (Ramstead, Constant, et al., 2019). The contribution to original knowledge of the second chapter is to propose a new approach to cognitive ecological dynamics, the variational ecology (VE), to answer the question of how to draw the boundaries of cognitive system beyond the organismic scale. VE embeds the systems studied by VNE (i.e., multiply nested, adaptively behaving cognitive systems) in a broader ecological context. This move was central to the development of a fully generalisable model of explanation in the cognitive sciences for studying multiscale systems of systems.

6.3. Moving beyond the debate between internalism and externalism: Rejection of essentialism about the boundaries of cognitive systems in the cognitive sciences

The fourth and final chapter of this thesis, "Multiscale integration: Beyond internalism and externalism," is published in *Synthese* (Ramstead, Kirchhoff, et al., 2019). It develops the philosophical implications of the theoretical construction of the first chapter and the technical results of second and third chapters, to provide a philosophical framework to ground and justify the multiscale, multidisciplinary frameworks exemplified by the first chapter – bringing my project full circle, as it were.

More specifically, building on the contributions of the previous chapters, the fourth chapter argues that the multiscale active inference framework can be used to dissolve a traditional distinction in the philosophy of the cognitive sciences – that between externalism and internalism. The aim of the final chapter is to show that the multiscale articulation of the FEP licences us to reject essentialist interpretations of the boundaries of cognition, and allows us to move beyond the debate between internalists and externalists.

Essentialism, here, is used in a specific technical sense – one that we have introduced in the literature via work in this paper. It is worth noting, of course, that our use is somewhat different from the usual sense of essentialism used in philosophy. Typically, an essentialist conception of some category or type of thing – e.g., philosophical types such as agency or intentionality – is one that associates with that type necessary and sufficient conditions, which determine whether some given thing should be classified as belonging to or being subsumed under that category.

In the third chapter, we apply the concept of essentialism to the debate between internalists and externalists in the philosophy of the cognitive sciences. This debate concerns the *boundaries of cognitive systems*; i.e., what counts as a cognitive system and what does not, and where to draw the relevant boundaries. Essentialism, on the reading we are employing here, is the position according to which cognitive systems are *defined essentially relative to some specific type of boundary*.

Internalist interpretations of the boundaries of cognition are those that claim that the only relevant factors for cognition are to be found inside the organism, and typically, they identify the most or only salient boundary as that of the skull. The philosopher Jakob Hohwy, for instance, a prominent defender of internalism, has claimed that the mind begins and ends at the spinal cord (Hohwy, 2014, 2016). Externalist positions endorse the view that cognition necessarily involves causal factors that extend beyond the boundary of the skull. These critics of the internalist view mainly operate under the umbrella of the '4E approach' to cognition

and ecological psychology. The philosophers Dan Hutto and Erik Myin, for instance, advocate an 'Equal Partners Principle' according to which the brain, the body, and the environment all contribute meaningfully and equally to the cognitive performances of living systems (Hutto & Myin, 2013). Though it is clearly more amenable to a multiscale view than internalism, most work on externalist approaches ends up endorsing an essentialist position.

Consider for instance the following characterisations of cognition from externalist philosophers. Proponents of embodied cognition make claims that sometimes sound essentialist, for instance when it is claimed that cognition "is not simply a brain event. It emerges from a process distributed across brain–body–environment… from a third person-perspective the organism–environment is taken as the explanatory unit" (Gallagher, 2017, p. 6). Proponents of the enactive approach make similar claims, for instance that cognition "is behavior or conduct in relation to meaning and norms that the system itself enacts or brings forth on the basis of its autonomy" (Thompson, 2010, p. 126); or that "there is not *a priori* reason to suppose that cognition is an exclusively heady affair [i.e., a processes realised within the brain]… cognition is fully embodied and embedded, and not merely embrained" (Hutto & Myin, 2013, p. 12). These seem like strong positions that define cognition essentially relative to a set of 'external' boundaries, i.e., those of the living body or those of the intentionally acting organism.

Against both internalism and externalism, the final chapter of this thesis proposes to reject all forms of essentialism. We leverage variational neuroethology and variational ecology to stake out a compromise position between (what we take to be) the overly coarse distinction, and false dichotomy, between internalism and externalism. We argue that the Markov blanket formalism licences a conception of the boundaries of cognition as: *nested*, in that most of the interesting cognitive systems studied in the cognitive sciences are composed of other cognitive systems; *multiple*, in that there will always be a plurality of relevant boundaries at play in any analysis of adaptive behaviour, namely, the boundaries of all the relevant nested Markov blanketed systems; and *interest-relative*, in that the most relevant boundary to study some cognitive process will vary as a function of the explanatory interests of cognitive scientists. This chapter is the first to do so via a discussion of the Markov blanket formalism that underwrites the FEP.

Furthermore, the model of explanation that is developed in this thesis does not merely tell us *that* the boundaries of cognitive systems are nested, multiple, and interest-relative; it provides us with a principled explanation of *why* the relevant boundaries are relevant. On the account presented here, a given boundary is relevant only to the extent that it mediates the

statistical dependencies between the system of interest and the systems with which it interacts. This allows us to determine which boundaries are the most pragmatically relevant to explain a given cognitive process. Since conditional independence is measurable, it effectively becomes possible to determine empirically which of boundaries are the most relevant, which provides a way of correcting internalist and externalist essentialism about the boundaries of cognitive systems.

The theoretical and mathematical models developed in the first three chapters were *guided* by work in philosophy, and also by theoretical work in philosophically-informed cognitive sciences. They leverage the FEP to show *that* and *how* one can work from a perspective that is able to produce bona fide models of multiscale phenomena; and provide a case study in multiscale modelling, applied to sociocultural cognition, action, and learning. Thus, in these chapters, it is primarily philosophical work that informs and extends the theoretical models that underwrite the FEP to phenomena beyond the brain. In the final chapter, the theoretical results obtained in the first three chapters feed back, as it were, into philosophical work in the narrower sense, where they are used to dispel the false dichotomy between internalism and externalism – and indeed, to dispel all essentialist conceptions of the boundaries of cognition.

6.4. Enactivism 2.0. or Bayesian enactivism

All three papers that form this thesis propose a new take on what has been called ‘pragmatist’ or ‘enactive’ cognitive sciences (Di Paolo & Thompson, 2014; Engel et al., 2016; Froese & Di Paolo, 2011; Thompson, 2010; Varela et al., 1991). Pragmatist or enactive cognitive sciences defend the claim that cognition is mostly (indeed, some might say almost exclusively) in the service of action; i.e., the claim that function of the brain is to guide contextually appropriate forms of adaptive behaviour.

As discussed above, the cardinal concepts of the enactive approach are *embodiment* and *enactment* (Thompson, 2010; Varela et al., 1991). Both of these ideas find their origin in the philosophical work of phenomenological philosophers that were interested in the living body, understood both as a physical thing that is part of the natural world – what German phenomenologists called *Körper* – and as a permanent feature of first-person phenomenological experience, the lived body or *Leib* (Husserl, 1990; Merleau-Ponty, 2013/1945). In brief, embodiment is the idea that the body and its dynamics in motion irreducibly feature in most (if not all) cognitive processes; as a physical thing, it structures the flow of perception and cognitive processes, which borrow from it its stability and structure;

and as a feature of conscious experience, it provides the raw material for the structuring of higher-order forms of thought.

Enactment is an idea drawn from pragmatism (James, 1975) and phenomenological philosophy (Heidegger, 2010/1927; Merleau-Ponty, 2013/1945), applied to the study of cognitive systems. Essentially, enactment is the process by which organisms actively make sense of their worlds through a *process* of interpretation or sense-making that irreducibly involves *embodied action in the world*. Meaning for organisms, on this view, is not some static thing that just needs to be 'picked up' by the organism; perception is not just a question of internally reconstructing a prespecified external world. Rather, meaning is something that is brought forth by the organism, through its active interpretation of the world and, crucially, through its patterns of adaptive action and perception. Thus, on this view, cognition is essentially the embodied activity of an organism. Much work in enactive theory has consisted in articulating the claim that seemingly disembodied activities like dreaming (Solomonova & Sha, 2016) and abstract or metaphorical thought (Barsalou, 1999, 2008) have their basis or are grounded in embodied activity.

The multiscale view proposed in this thesis critically reframes the enactive approach. Proponents of enactive cognitive sciences have had a very narrow understanding of information processing and of the place of computational methods in the cognitive sciences (Hutto & Myin, 2013; Thompson, 2010). This position was a productive stance at one point in the history of the cognitive sciences – i.e., during the late 1980s and 1990s, in the heyday of symbolic computational approaches to the study of cognition (Fodor, 1975; Fodor, 1983). Indeed, it led to an entirely new, *externalist* research paradigm for the cognitive sciences, which dominated the field from the 1990s onwards, to become one of the main approaches in the cognitive sciences today – what has become known as 4E approaches in the cognitive sciences (Clark, 2008; Clark & Chalmers, 1998; Kiverstein & Clark, 2009; Varela et al., 1991). These approaches largely rejected of the digital computer metaphor and the embrace of a picture of the brain as a dynamical system.

The series of papers in this thesis represents one of the first systematic attempts to bridge the more externalist perspectives of enactive-embodied-extended cognitive sciences and ecological psychology, and the originally more internalist leaning of the FEP. We propose a compromise position between internalism and externalism, essentially using tools developed in the context of a theory of the brain (the FEP and active inference) to advance an agenda compatible with the externalist emphasis on embodied activity and ecological embeddedness. The papers in this thesis, and especially the last three chapters, propose a

novel reading of embodiment and enactment that relies on their role in information processing – and thus reject the externalist move against information theory in the cognitive sciences. We claim that externalist theory has, as it were, thrown away the information baby with the computation and representation bathwater; and that the resources of information theory, computational modelling, and the FEP that emerges from its application to Markov blanketed systems, is – despite its internalist appearances – a great boon to enactive or pragmatist theories.

One of the contributions of the third chapter in this regard is to redefine affordances, providing ecological psychology with a mechanistic foundation based in theoretical biology. Crucially, the third paper extends the work done in the first paper by offering a new take on the construct of affordances, originally drawn from ecological psychology. Following recent work that syntheses ecological psychology and the enactive approach (Bruineberg & Rietveld, 2014; Chemero, 2009), as well as recent advances in the theory of active inference (Friston, Rigoli, et al., 2015; Parr & Friston, 2017), we redefine affordances via the FEP. Thanks to the framework provided by active inference, we distinguish the *action policy* itself – i.e., the possibilities for action that are available to the organism – and the *affordance* of that policy, which we interpret as the degree to which an organism is compelled to enact a policy. Under this reading, affordance is no longer construed as a possibility for action – that construct is instead resorbed under or subsumed by the notion of an action policy – but instead quantifies the extent to which an organism will be compelled to follow a given action policy. This gives us a mechanistic, quantitative grip on what the SIF calls 'solicitations'; i.e., those affordances that – 'here and now', in real-time interactions with the ecological niche – solicit the organism to act (Bruineberg & Rietveld, 2014). We quantify the affordance of a given action policy as the *negative expected free-energy* associated with a given action policy. This allows us to propose a *principle of most affordance* that is just another form of the principle of least action, and which follows directly from the FEP. This principle of most affordance is another contribution to original knowledge made by this thesis.

In summary, this thesis contributes to the emerging body of work that interprets the FEP and active inference through an enactivist or pragmatist lens. Most of the philosophical literature on active inference is centred on the role of the *brain* in perception and action (Clark, 2013, 2015b; Hohwy, 2014, 2016, 2017; Kiefer & Hohwy, 2017); the literature presents an active inference story that privileges causal factors internal to the organism in the explanation of its adaptive behaviour (neural dynamics) – and that does not foreground action or embodiment. The papers in this thesis make the claim that the active inference framework

is continuous with pragmatist and enactive cognitive sciences, building on work by Jelle Bruineberg in particular (Bruineberg, 2017; Bruineberg et al., 2016; Bruineberg & Rietveld, 2014). We show that active inference is a theory of action policy selection; it is about how organisms select the actions with the most adaptive value, which intrinsically lends itself to a pragmatist or enactivist reading. On our reading, organisms *embody* a 'best guess' (technically, a variational model) of the causes of their sensations; and they *enact* the relations captured in a normative (generative) model of their preferences, i.e., their phenotypical, preferred states.

7. Contribution of Authors

7.1. Chapter 1: "Cultural affordances: Scaffolding local worlds through shared intentionality and regimes of attention"

MJR did the initial conceptualization work, organised the argumentative structure of the paper, and wrote the initial draft of the manuscript, working in part from notes drafted by SV on cultural affordances. SV provided help with conceptualization and integration with cognitive and evolutionary anthropology. LK oversaw the writing process and contributed to the conceptualization of the paper.

7.2. Chapter 2: "Answering Schrödinger's question: A free-energy formulation."

MJR did the initial conceptualization work. MJR planned the manuscript and organised its argumentative structure, wrote the initial draft of the manuscript, and created the initial versions of some of the figures. PBB helped with the conceptualization. KJF helped with the theoretical articulation and formalization of nested systems within systems, notably providing the mathematical apparatus to do so.

7.3. Chapter 3: "Variational ecology and the physics of sentient systems"

MJR and AC did the conceptualization work. MJR and AC planned the manuscript, organised its argumentative structure, and wrote the initial draft of the manuscript. PBB helped with the conceptualization. KJF helped with the theory and formalization.

7.4. Chapter 4: "Multiscale integration: Beyond internalism and externalism"

MJR did the conceptualization work for the paper and identified the paper's main target, i.e., essentialism about the boundaries of cognition. MJR worked out the argument, planned the paper, and wrote the first draft of the manuscript. MDK helped with conceptualisation work and helped write the manuscript. AC helped with conceptualisation work as well and drafted parts of section 4. KJF ensured that the technical results that underwrite the paper were accurately presented.

8. Body of the thesis

8.1. Chapter 1: Cultural affordances: Scaffolding local worlds through shared intentionality and regimes of attention.

8.2. Chapter 2: Answering Schrödinger's question: A free-energy formulation.

8.3. Chapter 3: Variational ecology and the physics of sentient systems.

8.4. Chapter 4: Multiscale integration: Beyond internalism and externalism.

8.1. Chapter 1: Cultural affordances: Scaffolding local worlds through shared intentionality and regimes of attention

Original publication details:

Ramstead, M. J. D., Veissière, S. P. L., & Kirmayer, L. J. (2016). Cultural affordances: Scaffolding local worlds through shared intentionality and regimes of attention. *Frontiers in Psychology. Cognitive Science*, 7: 1090.
Research Topic: Beyond Embodied Cognition: Intentionality, Affordance, and Environmental Adaptation
doi.org/10.3389/fpsyg.2016.01090.

Authors:
Maxwell J. D. Ramstead[1,2*]
Samuel P. L. Veissière[2,3,4,5*]
Laurence J. Kirmayer[2*]

Affiliations:
[1]Department of Philosophy, McGill University, Montreal, Quebec, Canada.
[2]Division of Social and Transcultural Psychiatry, Department of Psychiatry, McGill University, Montreal, Quebec, Canada.
[3]Department of Anthropology, McGill University, Montreal, Quebec, Canada.
[4]Raz Lab in Cognitive Neuroscience, McGill University, Montreal, Quebec, Canada.
[5]Department of Communication & Media Studies, Faculty of Humanities, University of Johannesburg, Johannesburg, Gauteng, South-Africa.
* Corresponding authors

Abstract:
In this paper we outline a framework for the study of the mechanisms involved in the engagement of human agents with cultural affordances. Our aim is to better understand how culture and context interact with human biology to shape human behavior, cognition, and experience. We attempt to integrate several related approaches in the study of the embodied, cognitive, and affective substrates of sociality and culture and the sociocultural scaffolding of

experience. The integrative framework we propose bridges cognitive and social sciences to provide (i) an expanded concept of 'affordance' that extends to sociocultural forms of life, and (ii) a multilevel account of the socioculturally scaffolded forms of affordance learning and the transmission of affordances in patterned sociocultural practices and regimes of shared attention. This framework provides an account of how cultural content and normative practices are built on a foundation of contentless basic mental processes that acquire content through immersive participation of the agent in social practices that regulate joint attention and shared intentionality.

Keywords:
Affordances (ecological psychology); Cultural affordances; Radical embodied cognition; Enactive cognitive neuroscience; Free-energy principle; Predictive processing; Regimes of attention; Cognitive anthropology.

Acknowledgements:
Work on this chapter was supported by grants from the Social Sciences and Humanities Research Council of Canada (*Have We Lost Our Minds?*, M. J. D. Ramstead, award holder) and the Foundation for Psychocultural Research (*Integrating Ethnography and Neuroscience in Global Mental Health Research*, L.J. Kirmayer, PI). We thank Paul Badcock, Jelle Bruineberg, Paul Cisek, Karl Friston, Michael Kirchhoff, Frank Muttenzer, Kris Onishi, Ishan Walpola, Eric White, and two anonymous reviewers for helpful discussions and comments on earlier versions of this paper. Thanks to Marie-Ève Lacelle and Mariana Zarpellon for help designing our figures.

1. Introduction

The acquisition of culture is notoriously difficult to study. Over 70 years of research on the development of person-perception, for example, have made it clear that children as young as 4 years of age have already acquired implicit biases about ethnicity and other socially constructed categories of persons (Aboud & Amato, 2008; Clark, 1963; Clark & Clark, 1939; Hirschfeld, 1998; Huneman & Machery, 2015; Kelly, Faucher, & Machery, 2010; Machery & Faucher, 2005; Pauker, Williams, & Steele, 2016). These biases are consistent with the dominant culture of their societies, but are most often not consciously held or explicitly taught by their caregivers and educators. While most young children express a positive bias toward people they identify as members of their own group, children from minority groups typically show preferences for dominant groups, rather than for persons of their own ethnicity (Clark & Clark, 1939; Kinzler & Spelke, 2011). How such biases are acquired is still an open question. Ethnographic studies of socialization, education, and language acquisition have pointed to broad cross-cultural variations in how children are instructed, spoken to, expected to behave, involved in community activities, and exposed to other socializing agents beyond nuclear or extended families (Mead, 1975; Rogoff, 2003; Schieffelin & Ochs, 1986). However, by age 5, children across cultures have for the most part become proficient in the dominant set of expectations and representations of their cultures, despite the much discussed poverty of cultural stimuli to which they are exposed (Chomsky, 1965). These matters point to a human propensity for 'picking up' the broad scripts of culture even without any explicit instruction. In other words, we all come to acquire the shared background knowledge, conceptual frameworks, and dominant values of our culture. The presence of intuitive or implicit, yet stable and widely shared beliefs and attitudes among children constitutes a challenging problem for cognitive and social science.

In this paper, we outline a framework for the study of the mechanisms that mediate the acquisition of cultural knowledge, values, and practices in terms of perceptual and behavioural *affordances*. Our aim is to better understand how culture and context shape human behavior and experience by integrating several related approaches in the study of the embodied, cognitive, and affective substrates of action and the sociocultural scaffolding of embodied experience. The integrative framework we propose bridges cognitive and social sciences to provide (i) an expanded concept of 'affordance' that extends to sociocultural forms of life, and (ii) a multilevel account of the socioculturally scaffolded forms of affordance learning and the transmission of affordances in patterned sociocultural practices.

The context of the present discussion is the search for the 'natural origins of content' (Hutto & Satne, 2015). We hope to contribute to the naturalistic account of the emergence of semantic content, that is, of the evolution (in phylogeny) and acquisition (in ontogeny) of representational or propositional content. Cultural worlds seem to be full of meaningful 'content'—of explicit ways to think about and respond to the world in terms of kinds of agents, actions, and salient events. 'Content', here, is defined in terms of representational relations with satisfaction conditions: a vehicle *x* bears some semantic or representational content *y* just in case there are satisfaction conditions which, when they obtain, tell us that the vehicle is *about* something. Semantics is an intensional notion (Haugeland, 1990; Millikan, 1984, 2004, 2005; Piccinini, 2015). How do humans acquire this cultural knowledge and capacity to respond in social contexts in ways that actors and others find meaningful and appropriate?

We hypothesize that agents acquire semantic content through their immersion in, and dynamic engagement with, feedback or looping mechanisms that mediate shared intentionality and shared attention. Semantic content, we suggest, is realized in culturally shared expectations, which are embodied at various levels (in brain networks, cultural artifacts, and constructed environments) and are enacted in 'regimes' of shared attention. We generalize contemporary ecological, affordance-based models of cognitive systems adapting to their contexts over ontogeny and phylogeny to account for the acquisition of cultural meanings and for the elaborate scaffoldings constituted by constructed, 'designer' niches (Clark, 2015; Hutchins, 2014; Kirchhoff, 2015a). We suggest that 'regimes of shared attention'—that is, patterned cultural practices (Roepstorff, Niewöhner, & Beck, 2010) that direct the attention of participant agents—modulate the acquisition of culturally-specific sets of expectations. Recent work in computational neuroscience on predictive processing provides a model of how cultural affordances could scaffold the acquisition of socially shared representational content. In what follows, we shall sketch a multilevel framework that links neural computation, embodied experience, cultural affordances, and the social distribution of representations.

We begin by specifying a conceptual framework for 'cultural affordances', building on recent accounts of the notion of affordances in ecological, enactivist, and radical embodied cognitive science (Box 1). We propose to distinguish two kinds of cultural affordances: 'natural' affordances and 'conventional' affordances. Natural affordances are possibilities for action, the engagement with which depends on an organism or agent exploiting or leveraging reliable correlations in its environment with its set of abilities. For instance, given a human agent's bipedal phenotype and related ability to walk, an unpaved road affords a trek. Conventional affordances are possibilities for action, the engagement with which depends on

agents' skillfully leveraging explicit or implicit expectations, norms, conventions, and cooperative social practices. Engagement with these affordances requires that agents have the ability to correctly infer (implicitly or explicitly) the culturally specific sets of expectations in which they are immersed—expectations about how to interpret other agents, and the symbolically and linguistically mediated social world. Thus, a red light affords stopping not merely because red lights correlate with stopping behavior, but also because of shared (in this case, mostly explicit) norms, conventions, and rules. Both kinds of cultural affordances are relevant to understanding human social niches; and both natural and conventional affordances may be socially constructed, albeit in different ways (Hacking, 1999). Human biology is cultural biology; culture has roots in human biological capacities. The affordances with which human beings engage are cultural affordances.

We then assess the tensions between our proposed framework and radical enactivist and embodied approaches, which are typically committed to forms of non- (or even anti-) representationalism. On these views, perception, cognition, and action need not involve computational or representational resources. The scope of this claim varies. For some, this entails a rejection of computational or representational models and metaphors in the study of the mind—a staunch commitment to anti-representationalism (Chemero, 2009; Gallagher, 2001, 2008; Thompson, 2010; Varela, Thompson, & Rosch, 1991). More conciliatory positions instead suggest that basic cognitive processes are without content, but accommodate a place for contentful cognition. They claim that certain typically human forms of cognition involve representations, in the sense that human agents have the dispositions (mechanisms, behavioral repertoires, etc.) that are required to immersively engage with sociocultural content (e.g., patterned symbolic practices, linguistic constructions, storytelling and narration). We argue that contemporary computational neuroscience complements the more conciliatory of these approaches by providing a minimal neural-computational scaffolding for the skilled engagement of organisms with the available affordances.

Having done this, we turn to affordances in social and linguistic forms of life. We examine local ontologies, understood as sets of shared expectations, as well as the complex feedback relations (or looping effects) between these ontologies and human modes of communication, shared intentionality, and shared attention. Drawing on the skilled intentionality framework (Bruineberg & Rietveld, 2014), we examine the dynamics of cultural affordance acquisition through patterned cultural practices, notably attentional practices. We hypothesize that feedback mechanisms between patterned regimes of attention and shared forms of intentionality (notably shared expectations and immersion in local ontologies) leads

to the acquisition of such affordances. This framework can guide future research on multilevel, recursive, nested cultural affordances and the social norms and individual expectations on which they depend.

2. A theoretical framework for affordances

Much recent work in cognitive science has been influenced by the notion of affordances originally introduced by Gibson (1979). The interdisciplinary framework currently being developed to study affordances provides us with a point of departure for thinking about the evolution and acquisition of semantic, representational content. The aim of this section is to clarify the implications of adopting this framework.

Affordances are central to the emerging 'enactivist' and 'radical embodied' paradigms in cognitive neuroscience. Theorists of enactive cognition model the intelligent adaptive behavior of living cognitive systems as the dynamic constitution of meaning and salience in rolling cycles of perception and action, explicitly recognizing the emergence of meaning and salience in the active, embodied engagement of organisms with their environment (Di Paolo, 2009; Di Paolo, 2005; Di Paolo & Thompson, 2014; Froese & Di Paolo, 2011; Hutto & Satne, 2015; Hutto & Myin, 2013; Kirchhoff, 2016; Noë, 2004; Thompson, 2010). Embodied approaches in cognitive science explain the feats of intelligence displayed by cognitive systems by considering the dependence of cognition on the various aspects of the body as it engages with its environment, both internal and external (Barsalou, 2008; Shapiro, 2010). 'Radical embodied' cognitive science extends the theoretical framework of ecological psychology (Gibson, 1979) to the embodied cognition paradigm, providing a phenomenologically plausible account of active, dynamical coping (Bruineberg & Rietveld, 2014; Chemero, 2003, 2009; Rietveld & Kiverstein, 2014; Thompson & Varela, 2001). Recently, the enactive, radical enactive, and radical embodied approaches have been extended to 'higher-order' social and cultural systems (Froese & Di Paolo, 2011; Hutto & Myin, 2013; Rietveld & Kiverstein, 2014). This latter branch of enactivist theory will concern us especially.

2.1. Perspectives, affordances, and phenomenology

One of the distinctive contributions of ecological, radical embodied, and enactivist theories of cognition is their shared emphasis on the point of the view of the organism itself, understood as an intentional center of meaningful behavior. The implication of these 'perspectivist' approaches in cognitive science is that the world is disclosed as a set of 'affordances', that is, possibilities for action afforded to organisms by the things and creatures that populate its

environmental niche, as engaged through their perceptual and sensorimotor abilities (Heft, 2001; Silva, Garganta, Araújo, Davids, & Aguiar, 2013; Turvey, 1992; Turvey, Shaw, Reed, & Mace, 1981); cf. also Thompson (2010); Varela (1999). To paraphrase Wittgenstein, the world is the totality of possibilities of action, not of things. Perspectivist approaches in cognitive science operationalize this view of the organism and propose an account of perception, cognition, and action that is closer to the phenomenology of everyday experience.

Affordances provide an alternative framework for thinking about perception, cognition, and action that dissolves the strict conceptual boundary between these categories in a way that is closer to the phenomenology of everyday life.[1] This approach echoes the kernel insights of the phenomenology of Heidegger (2010/1927) and Merleau-Ponty (1968/1964, 2013/1945) about perception and action. Cognitive agents experience the world perceptually through the mediation of action, as a function of those actions that things in the world afford. For example, my cup of coffee is not first perceived as having such and such properties (size, shape, color), and only then as providing the opportunity for sipping dark roast. Instead, my filled cup is directly perceived as affording the action of sipping. Filled cups of coffee afford sipping; a paved road affords walking; a red traffic light affords stopping. The claim, then, is that cognitive agents typically do not encounter the world that they inhabit as a 'pre-given', objective, action-neutral set of things and properties, to be reconstructed in perception and cognition on the basis of sensory information, as classical models in cognitive science once suggested (e.g., Dawson, 2013; Fodor, 1975; Marr, 1982). The things that we engage are disclosed instead directly as opportunities for action—that is, as affordances. As Heidegger (2010/1927) famously argued, it is only when my smooth coping breaks down (say, when I run out of coffee, or when the cup breaks) that the objective properties of the cup become salient, present in perceptual experience at all.

---

[1] Enactive accounts reject the rigid separation of perception, cognition, and action, emphasizing that organisms cope with their environment in rolling cycles of engagement in which the distinction between action, cognition, and perception is blurred. When such a distinction is made, enactivist thinkers typically resist the traditional picture that subordinates action to perception or cognition. Theorists who draw the distinction nevertheless emphasize the deep connection between perception, cognition, and action. There are good reasons to think that action is a precondition for perception or that perception is a form of action (Clark, 2015; Kirchhoff, 2016). As we shall see in section 2., free-energy approaches frame perception and action as complementary ways of minimizing 'prediction error'. Our preference is to speak of rolling cycles of 'action-perception' to refer to the complex looping process whereby organisms cope with their environment. These cycles rely on various complementary computational strategies to minimize prediction error, which may (or may not) correspond to the traditional concepts of action, cognition, and perception.

The principal motivation for thinking of perception, cognition, and action in terms of engagement with affordances is that cognitive scientific accounts of these activities ought to be coherent with the phenomenology of action and perception in everyday life. Phenomenology tells us that there are dense interrelations between action and perception, that perception is mainly about the control of action, and that action serves to guide perception (Merleau-Ponty, 1968/1964, 2013/1945). Affordances provide a framework apt for this task, allowing us to integrate phenomenological experience into our models of explanation in cognitive science (Petitot, Varela, Pachoud, & Roy, 1999; Varela, 1996). As the story goes, in the wake of the behaviorist turn, experiential factors and mentalist language were banished from psychology (Skinner, 2011; Watson, 1913). Cognitive science rehabilitated mentalism, at least to some extent, in its postulation of cognitive states and processes (Fodor, 1975; Putnam, 1975). Most contemporary functionalist and mechanistic accounts of cognition, however, contend that it is possible to exhaustively explain a cognitive function by specifying its functional organization or the mechanism that implements that function (e.g., Bechtel, 2007; Craver, 2007). As we shall see presently, the perspectivist emphasis on the dynamics of the phenomenology of everyday life that characterizes enactive and ecological approaches allows us to account for cognitive functions with a conceptual framework that explicitly bridges the phenomenology of action and perception, system dynamics, and functionalist cognitive neuroscience.

2.2. Landscapes and fields

Affordances, as possibilities for action, are fundamentally interactional. Their existence depends both on the objective material features of the environment and on the abilities of different kinds of organisms. This dependence on interaction does not mean that affordances have no objective reality or generalizability (Chemero, 2003, 2009). Affordances exist independently of specific individual organisms. Their existence is relative to sets of abilities available to certain kinds of organisms in a given niche. ‘Abilities’, here, refers to organisms’ or agents’ capabilities to skillfully engage the environment, that is, to adaptively modulate its patterns of action-perception to couple adaptively to the environment. Without certain abilities, correlative opportunities for action are unavailable. Certain chimpanzees, for instance, are able to use rocks to cracks nuts. But for nuts and rocks to afford cracking, the chimp must already be cognitively and physiologically equipped for nut-cracking. In Chemero’s model of affordances, objectivity and subjectivity do not have separate ontological status; they co-exist and co-emerge relationally.

Building on Chemero (2003, 2009), Rietveld and Kiverstein (2014) define an affordance as a relation between a feature or aspect of organisms' material environment and an ability available in their form of life. 'Form of life' is a notion adapted from the later Wittgenstein (1953). A form of life is a set of behavioral patterns, relatively robust on socio-cultural or biographical time scales, which is characteristic of a group or population. We might say that each species (or subspecies), adapted as it is to a particular niche and endowed with specific adapted abilities, constitutes a unique form of life. Different human communities, societies, and cultures, with sometimes strikingly different styles of engagement with the material and social world, constitute different forms of life. There are thus at least two ways to change the affordances available to an organism: (i) by changing the material aspects of its environment (which may vary from small everyday changes in its architecture or configuration to thoroughgoing niche construction) and (ii) by altering its form of life or allowing it to learn new abilities already available in that form of life (interacting in new ways with an existing niche by acquiring new abilities through various forms of learning).

Following recent theorizing on affordances (Bruineberg & Rietveld, 2014; Rietveld, 2008a, 2008c; Rietveld & Kiverstein, 2014), we consider the distinction between the 'landscape' of affordances and the 'field' of relevant affordances. The claim is that, typically, organisms do not engage with one single affordance at a given time. The world we inhabit is instead disclosed as a matrix of differentially salient affordances with their own structure or configuration. The organism encounters the world that it inhabits as an ensemble of affordances, with which it dynamically copes and which it evaluates, often implicitly and automatically, for relevance. For an affordance to have 'relevance' here means that the affordance in question 'solicits' the individual, concrete organism by beckoning certain forms of perceptual-emotional appraisal and readiness to act. This occurs because affordances are both descriptive *and* prescriptive: *descriptive* because they constitute the privileged mode for the perceptual disclosure of aspects of the environment; and *prescriptive* because they specify the kinds of action and perception that are available, situationally appropriate and, in the case of social niches, expected by others.

The 'landscape' of affordances is the total ensemble of available affordances for a population in a given environment. This landscape corresponds to what evolutionary theorists in biology and anthropology call a 'niche' (Fuentes, 2014; Odling-Smee, Laland, & Feldman, 2003; Sterelny, 2007; Sterelny, 2015; Wilson & Clark, 2009). A niche is a position in an ecosystem that affords an organism the resources it needs to survive. At the same time, the niche plays a role vis-à-vis other organisms and their niches in constituting the ecosystem as a

whole. A typical ecosystem (that is, a physical environment where organisms can live) has multiple niches, which have some degree internal structure: affordances have a variety of dynamics relationships (one thing leads to another, depends on, reveals, hides, enables, other possibilities for action) (Pezzulo & Cisek, 2016). Thus, the niche is the entire set of affordances that are available, in a given environment at a given time, to organisms that take part in a given form of life. More narrowly, a niche comprises the affordances available to the group of organisms that occupy a particular place in the ecosystem—or, in the case of humans, the social world—associated with (and partly constituted by) a form of life.

The 'field' of affordances, on the other hand, relates to the dynamic coping and intelligent adaptivity of autonomous, individual organisms. The field refers to those affordances that actually engage the individual organism at a given time. Of those affordances available in the landscape, some take on special relevance as a function of the interests, concerns, and states of the organism. These relevant affordances constitute the field of affordances for each organism. They are experienced as 'solicitations', in that they solicit (further) affective appraisal and thereby prompt patterns of 'action readiness', that is, act as perceptual and affective prompts for the organism to act on the affordance (De Haan, Rietveld, Stokhof, & Denys, 2013; Frijda, 1986; Frijda, 2007; Rietveld, De Haan, & Denys, 2013). This engagement will vary in complexity, conformity, and creativity from pre-specified or pre-patterned ways of acting to "free" improvisation, as we shall see below.[2]

The field of affordances changes through cycles of perception and action. Changes in the situation that the organism engages give rise dynamically to different solicitations, as a function of the state of the organism, much the way a physical gauge field gives rise to different potentials as a function of the local forces (Sengupta, Tozzi, Cooray, Douglas, & Friston, 2016). Consider the action of drinking a cup of coffee. The filled cup affords a gradient (grasping, sipping), that is, a potential for coupled engagement. When generated by the organism-environment system, this gradient can be experienced by the organism as a

---

[2] Some might express unease at the mixed language we use, which straddles phenomenology, system dynamics, and cognitive functions. We take this as a virtue of the multilevel nature of the explanatory framework provided by the notion of affordances, which is operative at all of these different descriptive levels. Readers who would prefer to keep phenomenological description distinct from other explanatory levels (i.e., neural, social, cultural levels of explanation) can replace our talk of directly modulating the landscape or field of affordances with a more phenomenologically neutral concept, such as the organism's 'selective openness' (Bruineberg & Rietveld, 2014). With this terminology, we might say that changes in the patterns of activity in the organism (states, interests, etc.) and the environment shape the organism's selective openness to saliencies.

solicitation. The gradient is dissipated through engagement. The experience of satiation that follows drinking, combined with the fact that cup has been emptied, alter the field of affordances, which as indicated changes as a function of the states of organism and niche. Thus, the gradient is 'consumed' or dissipates after successful engagement.

2.3. Meaning and affordances

Not all affordances are of the same kind. Here we draw on Grice's theory of meaning to suggest an approach to the varieties of cultural affordances in terms of their dependence on *content-involving conventions*. We argue that the affordances in human niches (what we call generally 'cultural' affordances) are of two distinct kinds: 'natural' and 'conventional' affordances.

Grice's theory of meaning, elaborated in a series of papers in the philosophy of mind (Grice, 1957, 1969, 1971, 1989), and later refined by Sperber and Wilson (1986), Levinson (2000), and Tomasello (2014), is often termed 'intention-based semantics', or 'implicature'. On a Grician account, meaning lies in a speaker's communicative intent; that is, in what she intends to convey through an utterance. Grice elaborated the first formula of his theory of meaning in these terms (using the subscript $_{NN}$ to signify to 'non-natural'):

> "A meant_{NN} something by X" is roughly equivalent to "A uttered X with the intention of inducing a belief by means of the recognition of this intention" (Grice, 1989, p. 19)

Taking this model beyond the dyadic sphere of conversational implicature, Grice later attempted to explain how "timeless" (that is to say, durable and widely shared) conventions of meaning are recognized in a shared cultural repertoire:

> "x means_{NN} (timeless) that so-and-so" might at a first shot be equated with some statement or disjunction of statements about what "people" (vague) intend (with qualifications about "recognition") to effect by x (Grice, 1989, p. 220)

In the subsequent 'relevance' account, Sperber & Wilson (1986) translated this automatic 'first shot' recognition of conventional meaning as one in which human minds scan for salient, meaning-generating cues in the environment, and stop processing when the cues are secured.

Our model draws on Grice to describe the stabilization of cultural cues as affordances. Key to our approach is the implied ontological and epistemic status of other minds (that is, the intentions of 'persons') in the embodied cognitive work required in the 'recognition', or more precisely, the enactment of meaning. Our proposal, then, is to follow Grice in understanding the thought, affect, and behavior of human agents as determined by implicit expectations about

others' expectations. Specifically, we argue that *humans behave according to the way they expect others to expect them to behave* in a given situation (see Figure 1).[3] As we shall explicate below, we contend that humans operate (often pre-reflectively) within the landscape and field of possibilities for *variations* in action[4] as a function of their expectations about what others expect of them in specific contexts (see Figure 2).

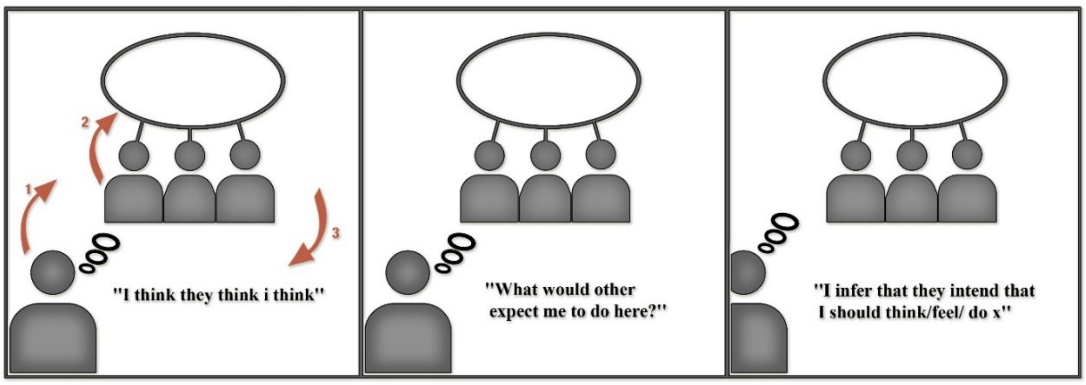


Fig. 1. *Basic cognitive formula*. Three orders of automatic intentionality.

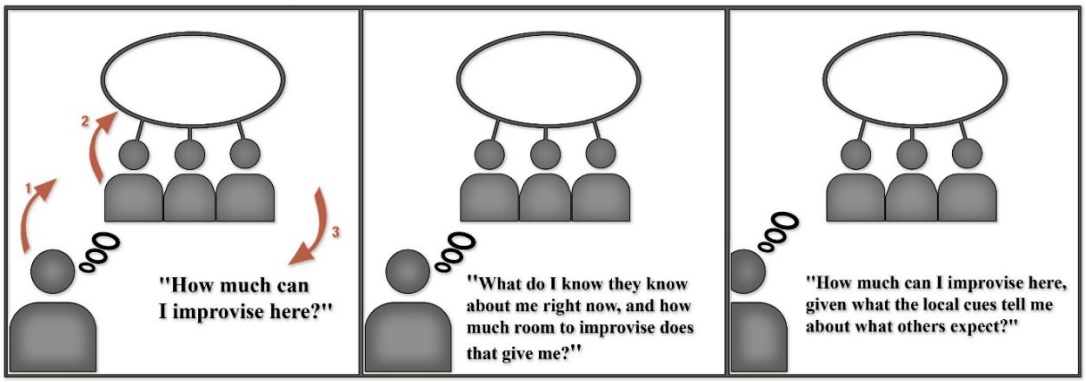


Fig. 2. *Full cognitive formula*. Three orders of intentionality governing improvisational variations in action.

---

[3] This basic cognitive formula for sociality requires three orders of automatic intentionality; that it is to say, an implicit, non-narrative, hypothesis-generating, error-reduction scenario of the "I think they think I think" variety that can be translated as "what would relevant others expect me to think/feel/do in this situation?"

[4] The notion of *variation* and improvisation in action within a convention is very important. Humans do not simply obey prescribed expectations, but also resist, transgress and transform them. Specific fields of joint-intentional affordances, thus, invariably entail different licenses for improvisation on expected behavior. The background formula for action is not simply "what would others expect me to do here?" but also "how much license or room to improvise do I have here given what the set of local cues tells me about others' expectations and the norms that should otherwise govern my behavior in this specific situation?"

The importance of these revisions to Grice's model of meaning to our framework for cultural affordances is to highlight the dependence of certain kinds of affordances on joint intentionality, and effective social and cultural normativity and conventionality, or equivalently, the shared expectations (both implicit and explicit) that codetermine the affordance landscape and local field dynamics. Grice (1957) distinguished between natural and non-natural forms of meaning, emphasizing the latter in most of his work. Natural meaning is a relation between two things that are correlated. Smoke 'means' fire because tokens of smoke reliably correlate with tokens of fire. Similarly, (certain kinds of) spots mean measles (understood not as the popular category but as the biomedically recognized infection with a particular virus). Non-natural meaning instead depends on the capacity of individual agents to exploit explicit and implicit social 'conventions' (in the wide sense of locally shared norms, values and moral frames, expectations, ontologies, etc.) to infer the intentional states of other agents and thereby engage them or engage aspects of the environment with them. Red traffic lights, in virtue of convention (and law), 'mean' stop, and hence afford (and mandate) stopping—and this is made possible by the specifically human mastery of recursive inferences, both explicit and implicit, that agents make about other agents (Tomasello, 2014).

Recent work on information processing has extended Grice's framework to account for different kinds of *information* (Piccinini, 2015; Piccinini & Scarantino, 2011; Scarantino & Piccinini, 2010). A token informational vehicle *x* of kind *X* (that is, a sign, a pattern of neural activation, or what have you) carries 'natural information' about some information source *y* of kind *Y* just in case there are reliable correlations between *X* and *Y*. Natural information, in other words, cannot misrepresent, for it is non-semantic; it is not the kind of thing that can be simply true or false. Such information can be exploited and leveraged by a cognitive system to guide intelligent behavior. Conversely, 'non-natural information' (or as we prefer to put it, 'conventional information'), pertains to semantic, content-involving representations that depend on social norms and cultural background knowledge. Non-natural information allows an agent to make a correct inference about some aspect of an intentional system, e.g., other agents, language and other symbolic systems such as mathematics, etc. Non-natural information is semantic in that it obtains in virtue of satisfaction conditions (e.g., truth conditions). A vehicle carries this kind of information about some state of affairs just in case some (explicit or implicit) shared convention, in the sense outlined above, links a vehicle to what it represents.

In the psychological and anthropological literature, affordances are usually understood as interactional properties between organisms and their environment that can be individually discovered in ontogeny without social learning. Chimpanzees, for example, rediscover how to crack nuts with rocks in each generation without vertical social transmission of skills (Howes, 2011; Ingold, 2000, 2001; Moore, 2013). Most of what humans do, in contrast, is learned socially and requires complex forms of coordination. We suggest, however, that successfully learned human conventions that govern action are also best conceptualized as affordances. Such affordances depend on *shared sets of expectations*, reflected in the ability to engage immersively in patterned cultural practices, which reference, depend on, or enact folk ontologies, moralities and epistemologies. We might call these '*conventional' affordances*.

An empty street affords being walked on or driven on to the lone pedestrian or driver. Yet affordances, especially those depending on conventions, might differ depending on context. A red traffic light, as we have seen, affords an agent stopping, particularly in the presence of others, and especially in the presence (real or imagined) of police who are expect to intervene. But a driver might alter her behavior as a result of not being seen by others. A red traffic light in an empty street at 4:00 AM, thus, might afford transgression of the stopping rule following an inference about the absence of other minds likely to judge the agent. Departing from Grice and earlier theories of information processing (Dretske, 1995), one might understand the notion of information as probabilistic: to carry information implies only the truth of a probabilistic claim (Scarantino, 2015; Scarantino & Piccinini, 2010). Although this account was developed for natural information, we extend it here to conventional information, given the prominence of social improvisation. 'Conventions' need not be explicitly formulated as rules, and may instead originate in the actors' engagement with local backgrounds over time, that is, from non-contentful developmental experiences, learning, or participation in social and cultural practices (Piccinini, 2015; Satne, 2015).

A cultural artifact may have multiple affordances according to its embedding in larger webs of relationships that are part of the individual's history of learning and the expectations for the potential participation of others. Indeed, to operate with conventional affordances, agents must have shared sets of expectations—we must know what others *expect us to expect* Simple rule-governed models of sociality go on the assumption that conventions lead to stable, binary affordances, where satisfaction conditions are either met or not. However, cultural symbols and signs are usually polysemous and their interpretation depends on context. Moreover, variations in the way agents engage with affordances in practice, often license what we could term 'skilled improvisation'. Rules and conventions can be followed slavishly,

selectively ignored, deliberately transgressed, or re-interpreted to afford new possibilities. Natural dispositions for shared intentionality in what Searle (Searle, 1991, 1992, 1995, 2010) calls the deep background, on this view, give rise to cooperative action not only through convention but also through iterative variations governed by modes of engagement with cultural affordances (Terrone & Tagliafico, 2014).

3. The neurodynamics of affordances

Some aspects of culture clearly involve content in the improvisational sense of the term: namely, those affordances that depend on conventions, social normativity, and the ability to improvise from a joint-intentional background enriched by cultural learning. Here, we aim to contribute to the effort to explicate the mechanisms by which basic minds are scaffolded into more elaborate content-involving processes. To explain agents' engagement with contentful affordances requires a theory of cultural content and representations.

Our hypothesis, to be explicated below, is that feedback loops mediating shared attention and shared intentionality are the principal mechanism whereby cultural (especially conventional) affordances are acquired. Before proceeding, however, we must face an objection stemming from tensions between our enactivist-embodied-ecological framework and our aim of providing a theory for the acquisition of semantic content. We have suggested that conventional affordances depend on shared expectations, perspective-taking, and even mindreading abilities. However, proponents of radical embodiment and enactivism argue that cognition can be understood as the coupling of an organism to its niche through dynamical processes, without any need to invoke representational processes and resources like explicit expectations and mindreading (Chemero, 2009; Gallagher, 2001, 2008; Thompson, 2010; Varela et al., 1991). On these accounts, classical theories of cognition (Fodor, 1975; Marr, 1982), which modeled cognition as the rule-governed manipulation of internal representations, radically misconstrue the nature of agents' intentional engagement with their worlds. The claim, then, is that much cognition can (indeed, must) be explained by appealing only to dynamical coupling between organism and environment.

Rejecting the claim that cognition necessarily involves representations, radical enactivists insist that basic cognitive processes ('basic minds') can function entirely without content (Hutto & Myin, 2013; Thompson, 2010). The argument, then, is that minds, especially basic minds like those of simple organisms (and many of the unreflective embodied engagements of more complex minds), do not require content. They only require adequate forms of coupling, which need bear no content at all. Adequate coupling only requires an

organism to leverage correlations that are reliable enough to be exploited for survival. This poses a challenge to a theory like ours, which aims to explicate the acquisition of cultural content in the form of conventional affordances. In this section, we accommodate this radical minimalism about representations and semantic content while sketching a neural computational account of the scaffolding of cultural affordances.

3.1. Computation, representation, and minimal neural models

Recent work on computation and neurodynamics helps to clarify the scope of radical arguments against content-involving, representational theories of cognition. Although older semantic theories view computation as the processing of representations (with propositional content and satisfaction conditions) more recent theories do not make this assumption. The 'modeling view' of computation (Chirimuuta, 2014; Grush, 2001; Shagrir, 2006, 2010) suggests that computation in physical systems (calculators, digital and analog computers, neural networks) employs a special kind of minimal, structural or analogical model based on statistical correlations (O'Brien & Opie, 2004; O'Brien & Opie, 2009, 2015). On this view, a computational process is one that dynamically generates and uses a statistical model of a target domain (say, things in the visual field). The model is said to 'represent' that domain only in the sense that the relations between its computational vehicles (digits, neural activation patterns, or what have you) preserve the higher-order statistical, structural-relational properties of the target domain, which can be leveraged to guide adaptive action. We might call this 'weak' (non-propositional) content, based on structural analogy between vehicle and target domain (O'Brien & Opie, 2004, 2009, 2015). Such statistical models are much more minimalistic than traditional representational theories of mind, which require that internal representations bear propositional content (Fodor, 1975). Even more minimalistic accounts of computation are available. Computation can be defined mechanistically, as the rule-governed manipulation of computational (rather than representational) vehicles (Miłkowski, 2013; Piccinini, 2015). On the mechanistic account, computations (digital, analog, neural) can occur without *any* form of semantic content (Piccinini & Scarantino, 2011; Scarantino & Piccinini, 2010).

Thus, some of the newest theories of computation are minimalistic about the representational nature of neural processes. Whether the modeling-structural and the mechanistic minimal statistical models deserve the label 'representation' is debatable (Anderson & Chemero, 2013; Clark, 2015; Hutto & Satne, 2015). To some degree the conflict may be merely terminological. What matters for our purposes is to note that the minimalistic

statistical-computational models in the cognitive system can be leveraged to guide skilled intelligent, context-sensitive, adaptive behavior. This provides additional weight to the claim that basic minds are without strong, propositional, semantic content (Hutto, Kirchhoff, & Myin, 2014; Hutto & Myin, 2013).

While this may be the case, human societies clearly transact in content-laden representations. We use language replete with images, metaphors and other symbols to tell stories and narrate our lives. We imagine particular scenarios or events, and we think about, describe, elaborate and manipulate these images or models in ways that treat them as pictures or representations of possible realities. Importantly, even on the radical view on offer here, nothing *precludes* such content-involving cognition. In recent discussions around the natural origins of content, it is hypothesized that neural computations can come to acquire representational content when coupled adequately to a niche or milieu through dense histories of causal coupling (Hutto & Satne, 2015; Hutto & Myin, 2013; Kirmayer & Ramstead, 2017). We suggest that immersive involvement of agents in patterned cultural practices during development, and the subsequent practice of the abilities acquired in enculturation, allows for the acquisition of stable cultural affordances. In the case of human beings, whose learning is mostly social, the function of the neural computations performed by a system becomes that of interfacing adequately with both representational and non-representational aspects of culture so as to guide appropriate behavior.

3.2. Free-energy and the neurodynamics of affordances

The framework we think can account for the acquisition of cultural affordances by agents rests on recent work in computational neuroscience and theoretical biology on the 'free-energy principle'. The free-energy principle is a mathematical formulation of the tendency of autonomous living systems to adaptively resist entropic disintegration (Friston, 2012a; Friston, 2013a; Friston, 2013b; Friston, 2010; Friston, Kilner, & Harrison, 2006; Sengupta et al., 2016). This disintegration can be thought of as the natural tendency of all organized systems (which are by their nature far-from-equilibrium systems) to dissipate, that is, to return to a state of low organization and high entropy or disorder—in other words, to return to (thermodynamic) equilibrium. The free-energy principle states that the dynamics of living organisms are organized to maintain their existence by minimizing the information-theoretic quantity 'variational free-energy'. By minimizing free-energy, the organism resists entropic dissipation and maintains itself in its phenotypical steady-state, far from thermodynamic equilibrium (death).

One application of the free-energy principle in computational neuroscience is a family of models collectively referred to as 'hierarchical predictive processing' models, which instantiate a more general view of the brain as a 'prediction machine' (Bar, 2011; Clark, 2015; Friston, 2012b; Friston & Kiebel, 2009; Friston, 2010, 2011; Frith, 2007; Hohwy, 2014) – for empirical evidence see Adams, Bauer, Pinotsis, & Friston (2016). In this framework, the brain is modeled as a complex dynamical system, the main function of which is to 'infer' (in a qualified sense) the distal causes of its sensory stimulation, starting only from its own sensory channels. The strategy employed by the brain, according to this view, is to use a 'generative model' of the distal causes and engage in self-prediction (Eliasmith, 2005; Friston, 2010). That is, the system's function is to predict the upcoming sensory state and compare it the actual sensory state, while minimizing the difference between these two distributions (predictions and prediction errors) through ongoing modification of predictions or action on the environment (see Figure 3 and Figure 4).

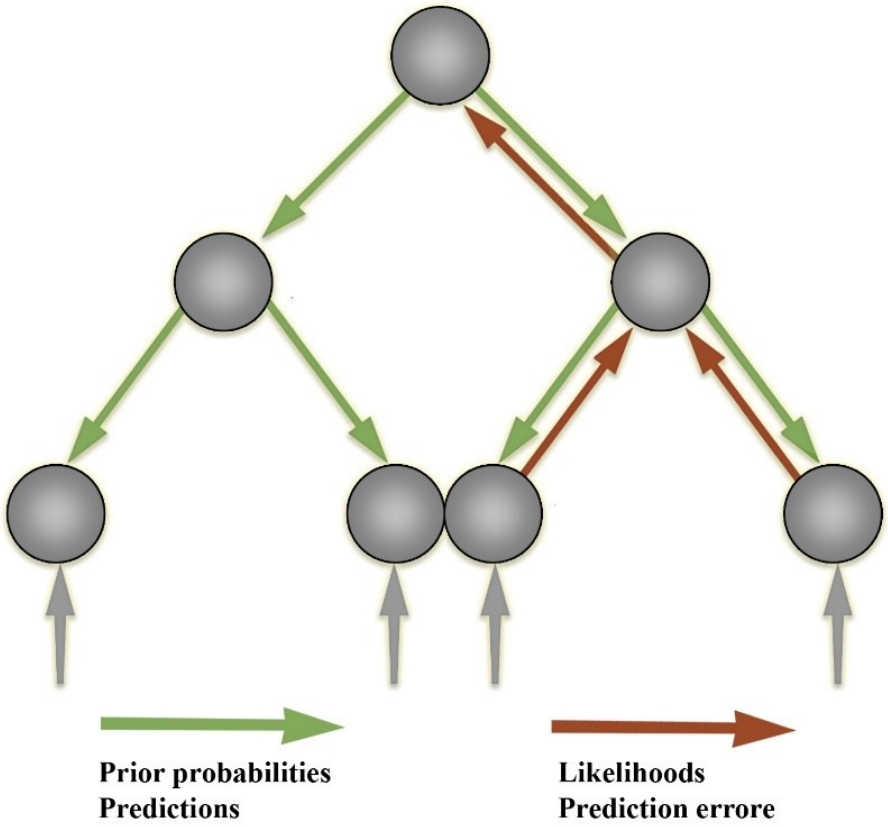


Figure 3. In the predictive processing approach, the main activity of the nervous system is to predict upcoming sensory states and minimize the discrepancy between prediction and sensory states ('prediction errors'). The information propagated upward to higher levels for further processing consists only in these prediction errors.

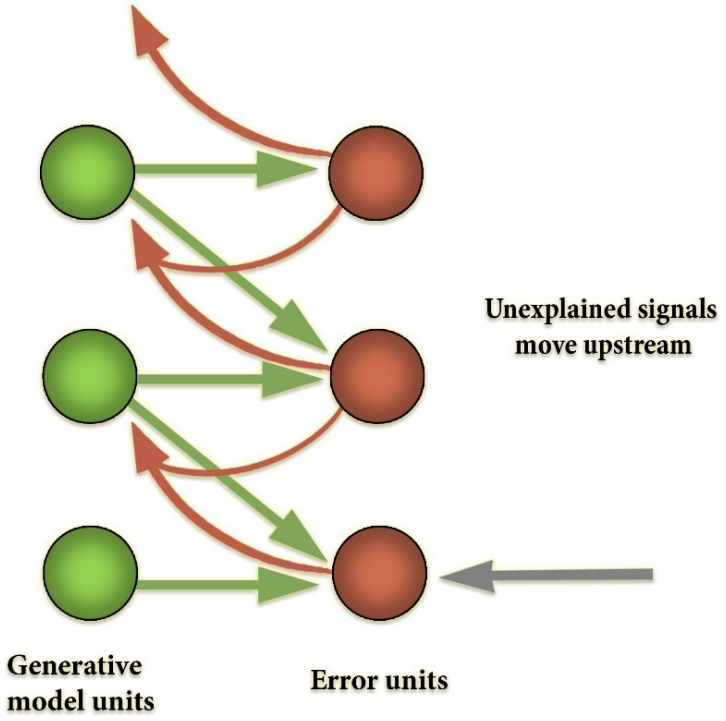


Fig. 4. *A diagram of Bayesian inference in predictive processing architectures.* The dynamics of such systems conform to the principles of the Bayesian statistical inference framework. The Bayesian statistical framework is central to predictive processing architectures, for the latter assume that neural network interactions operate in a way that maximizes Bayesian model evidence. Bayesian methods allow one to calculate the probability of an event taking place by combining the 'prior probability' of this event (the probability that such an event takes place before considering any evidence) with the 'likelihood' of that event, that is, the probability of that event given some evidence. This allows the Bayesian system to calculate the 'posterior probability' of the event, that is, the revised probability given any new available evidence. Prior probabilities are carried by predictions (green arrows) issued by the generative model units (green units). Likelihoods are carried by prediction errors (red arrows) issued by the error units (red units). In the 'empirical Bayes' framework, the system can then use the posterior obtained from one iteration as the prior in the next iteration. Predictions issued from the generative models, which encode prior beliefs, propagate up, down, and across the hierarchy (through backwards and lateral connections) and are leveraged to guide intelligent adaptive action-perception. This leveraging is achieved by cancelling out (or 'explaining away') discrepancies, which encode likelihood, through rolling cycles of action-perception. This same process allows the system to learn through plastic synaptic connections, which are continuously updated through free-energy minimization in action-perception. The system thus continuously and autonomously updates its 'expectations' (Bayesian prior beliefs) in rolling cycles of action-perception.

‘Generative models’ are minimal statistical models, of the kind discussed above. The use by a system of generative models need not entail semantic content. Their function is to dynamically extract and encode information about the distal environment as sets of probability distributions. The information involved here can be natural or conventional in kind. The only entailment is that the system or organism must leverage its generative model to guide skilled intentional coupling. The system uses this generative model to guide adaptive and intelligent behavior by ‘inverting’ that model through Bayesian forms of (computational, subpersonal) inference, allowing it to leverage the probability distributions encoded in the model to determine the most probable distal causes of that distribution and to act in the most contextually appropriate way (Clark, 2015; Friston, 2010; Hohwy, 2014).

How does this inversion take place? Generative models are used to generate a prediction about the upcoming sensory distribution. Between the predicted and actual sensory distributions, there almost always will be a discrepancy (‘prediction error’), which ‘tracks’ surprisal (in the sense that, mathematically, it is an upper bound on that quantity). The free-energy principle states that all living systems act to reduce prediction error (and thereby implicitly resist the entropic tendency towards thermodynamic equilibrium—dissipation and death). This can occur in one of two complementary ways: (i) through action, where the best action most efficiently minimizes free-energy by making the world more like the prediction (‘active inference’); and (ii) through perception and learning, by selecting the ‘hypothesis’ (or prediction, which corresponds to the probable distal cause of sensory distribution) that most minimizes error, or changing the hypotheses when none fits or when one fits better (Friston, 2013a; Friston & Frith, 2015a; Friston, 2011; Friston & Frith, 2015b; Friston et al., 2012). Given that generative models embody fine-grained statistical information about the distal environment at different scales, the top-down prediction signals (produced by higher levels in the processing system) provide crucial contextualizing information for the activity of lower levels in the predictive hierarchy, rendering the feedforward error signal contextually sensitive and adaptive (see Figure 5).

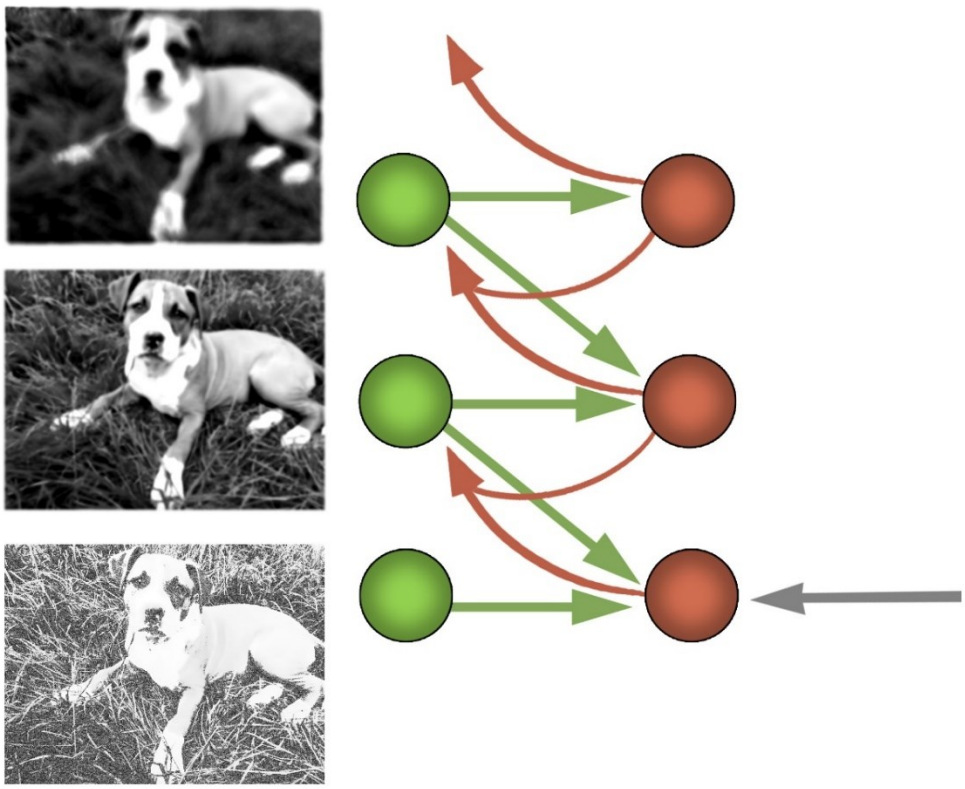


Fig. 5. *Diagram of hierarchical structure of the predictive processing networks.* Predictive networks have hierarchical structure in the sense that their processing is layered. The layered (hierarchical) structure of the generative model allows the model to capture the nested structure of statistical regularities in the world. This inferential architecture effectively allows the system to leverage new information dynamically and implement a 'bootstrapping' process, whereby the system extracts its own priors from its dynamic interactions with the environment. Computationally, each individual layer has the function of extracting and processing information leveraged to cope with regularities at a given level or scale. In this example, information about the visual scene is decomposed into high, medium, and low spatial frequency bands. Typically, low spatial frequency features change at a faster than high spatial frequency features. As such, lower spatial frequency information is encoded higher up in the processing hierarchy, to guide lower-level, faster processing of higher spatial frequency information. The hierarchical or layered statistical structure of the generative model enables it to recapitulate the salient statistical structure of those systems to which it is coupled. As discussed in the text, this need not imply semantic content (but does not exclude it either).

The representational minimalism of embodied generative models nicely complements the representation-sparse phenomenology of affordances. Such minimal models might be described as exploiting (non-semantic) information *for* affordances, rather than (semantic) information *about* affordances (van Dijk, Withagen, & Bongers, 2015); that is, the sensory

array only carries information given certain uses of it by organisms (i.e., being a statistical proxy). The 'internal representations' involved here might best be thought of as transiently 'soft-assembled neural ensembles', adequately coupled to environmental affordances (Anderson, 2014).

It can be argued that predictive processing models complement enactivist and radical embodied approaches and are compatible with minimalism about representations, provided we do not interpret the statistical computations and error signal processes in a strong semantic, content-involving sense (Hutto & Satne, 2015; Kirchhoff, 2015a, 2015b, 2016; Kirmayer & Ramstead, 2017). Generative models are simply *embodied statistical models* that are dynamically leveraged to guide intelligent adaptive behavior.

Generative models are embodied at different systemic levels and timescales, in different ways. As indicated, at the level of the brain, the predictive hierarchical architecture of neural networks come to encode statistical regularities about the niche, which allow the organism to engage with the field of affordances in adaptive cycles of action-perception. But the embodiment of generative models does not stop at the brain. Indeed, one radical implication of the free-energy principle is that the *organism itself is* a statistical model of its niche (Friston, 2013b; Friston, 2011). States of the organism (i.e., its phenotype, behavioral patterns, and so forth) come to statistically model the niche that it inhabits over evolutionary timescales (Badcock, 2012). Thus, phylogeny conforms to the free-energy principle as well, because the effects of natural selection is to select against organisms that are poor models of their environments. Those organisms that survive and thrive are those that embody, in this literal sense, the best generative models of their niche. Organism phenotypes can be described as conforming to the free-energy principle over developmental timescales in morphogenesis as well (Friston, Levin, Sengupta, & Pezzulo, 2015). Generative models are thus not only 'embrained', but embodied in an even stronger sense, over the timescales of phylogeny and ontogeny. This strong embodiment allows one to interpret free-energy approaches in a non-internalist way and to counter some objections raised against earlier formulations of predicting processing approaches (e.g., Clark, 2013; Hohwy, 2014). This multilevel embodiment of the generative model, as we shall argue below, extends to the concrete, material, human-designed milieus (or 'designer environments') in which humans operate.

Some generative models (in this wide sense) involve semantic content and others do not (they involve something more minimal than satisfaction conditions, i.e., reliable covariation). The study of minds without content is compatible with more extensively content

involving forms of (social and cultural) cognition that are scaffolded on such basic minds through processes of social learning and enculturation.

On the radical enactivist account, content-involving forms of intentionality emerge in the context of certain cultural practices in human forms of life (Hutto & Satne, 2015). Many of these practices involve multi-agent situations in which proper engagement requires forms of implicit perspective-taking and perspective-sharing (Sterelny, 2015). In some cases, such practices can involve explicit 'mindreading' as well, that is, inferring the beliefs, intentions, and desires of other agents *as such* (Michael, Christensen, & Overgaard, 2014). There is a long-running debate among anthropologists over the extent to which inferences about other people's mental states (as opposed to, say, bodily states) may reflect a folk psychology that is more pronounced among modern Western peoples (Robbins & Rumsey, 2008; Rumsey, 2013). This 'transparency of mind' folk psychology is contrasted in the literature with so-called 'opacity doctrines' found in other cultures, in which people's interior states are said to be 'opaque', or unknowable. As recent multi-systems account of social cognition have shown, however, situations involving novel cues or too many orders of intentionality will often trigger 'higher' cognitive resources and compel humans to think about other people's intentions as such (Michael et al., 2014). Engagement with affordances in the human niche also often requires 'mindshaping', as our interpretation of other agents' intentional profiles in turn shapes those same profiles through interpersonal loops (Sterelny, 2007; Sterelny, 2015; Zawidzki, 2013). Perspective-taking can be implicit and embodied in that organisms can act on situations by leveraging minimal models that encode information about other agents and their behavior without entailing the presence of semantic content (i.e., having satisfaction conditions). But this is not incompatible with the claim that perspective-taking and mindshaping abilities, in the human niche, often involve symbolically and linguistically mediated forms of communication, which substantially change the kind of affordance landscape available to human agents (Kiverstein & Rietveld, 2013; Rietveld & Kiverstein, 2014).

Although the perspectivist focus on the dynamic embodied enactment of meaning in a shared social world is central to our understanding of cultural affordances (Fuchs & De Jaegher, 2009; Gallagher, 2001, 2008), our contention is that the acquisition of representational content in 'epidemics' of socially shared representations (Claidière, Scott-Phillips, & Sperber, 2014; Sperber, 1996) entails that cognitive agents must be endowed with a neural-computational scaffolding adequate to such activities.[5] Even though basic cognition (and indeed, some forms

[5] We should note a few limitations of the 'epidemic' metaphor: (i) representations are not

of 'higher' cognition; Hutto & Myin, 2013) may be without content, given the symbolic and linguistic nature of human experience and culture, the human cognitive system must be equipped with the neural-computational resources needed to adequately couple with shared social representations, if we are to account for how the latter are transmitted stably and reliably. Semantic content is acquired through dense histories of embodied engagement with the environment. For humans, this involves participation in patterned, linguistically and symbolically mediated practices—which include patterns of shared attention and shared intentionality.

3.3. Predictive processing and attention

One aspect of the architecture of predictive processing is crucial for our account of cultural affordances: The predictive processing model specifies a deep functional role for *attention*. Attention, on the predictive processing account, is modeled as 'precision-weighting', that is, the selective sampling of high precision sensory data, i.e., prediction error with a high signal-to-noise ratio (Feldman & Friston, 2010). The efforts of the cognitive system to minimize free-energy operate not only on first-order, correlational statistical information about the distal environment, but on second-order statistical information about the signal-to-noise ratio or 'precision' (that is, inverse variance) of the prediction error signal as well. This allows the system to give greater weight to less noisy signals that may provide more reliable information. Based on this information, the cognitive system balances the gain (or 'volume') on the units carrying prediction errors at specific levels of the hierarchy, as a function of precision. This control function, in effect, controls the influence of encoded prior beliefs on action-perception (Friston, 2010). Greater precision means less uncertainty; the system thus 'ups the volume' on high precision error signals to leverage that information to guide behavior. Attention, then, is the process whereby synaptic gain is optimized to 'represent' (in the sense of reliably co-varying with) the precision of prediction error in hierarchical inference (Clark, 2015; Feldman & Friston, 2010).

Precision-weighting is centrally important in these architectures and has been proposed as a mechanism of *neural gating*. Gating is the process whereby effective connectivity in the brain (Friston, 1994, 2011), that is, the causal influence of some neural units on others, is

---

merely transmitted through contagion, but through many different means, modes of communication, and practices that are themselves culturally mediated; (ii) they reside not just in individuals, but also in artifacts and institutions; and (iii) they are usually not simply replicated, but modified or transformed by each individual or institution that takes them up.

controlled by the functioning of distinct control units (Daw, Niv, & Dayan, 2005; den Ouden, Daunizeau, Roiser, Friston, & Stephan, 2010; Stephan et al., 2008). These are called 'neural control structures' by Clark (1998). (For assessments of the empirical evidence, see: Friston, Bastos, Pinotsis, & Litvak, 2015; Kok, Brouwer, van Gerven, & de Lange, 2013; Kok, Jehee, & De Lange, 2012) Attention-modulated 'gating' is the central mechanism that allows for the formation of transient task- and context-dependent coalitions or ensembles of neural units and networks (Anderson, 2014; Park & Friston, 2013; Sporns, 2010).

Thus, in the predictive processing framework, attention is the main driver of action-perception. Clark (2015, p. 148ff.) describes possible implementations of this scheme in the brain. Much like for first-order expectations, the system encodes expectations about precision in the generative model, presumably in the higher levels of the cortical hierarchy (Friston, Stephan, Montague, & Dolan, 2014). These signals, which carry context-sensitive second-order statistical information, then guide the balancing act between top-down prediction signals from the generative models and bottom-up error signals in attention (see Figure 6).

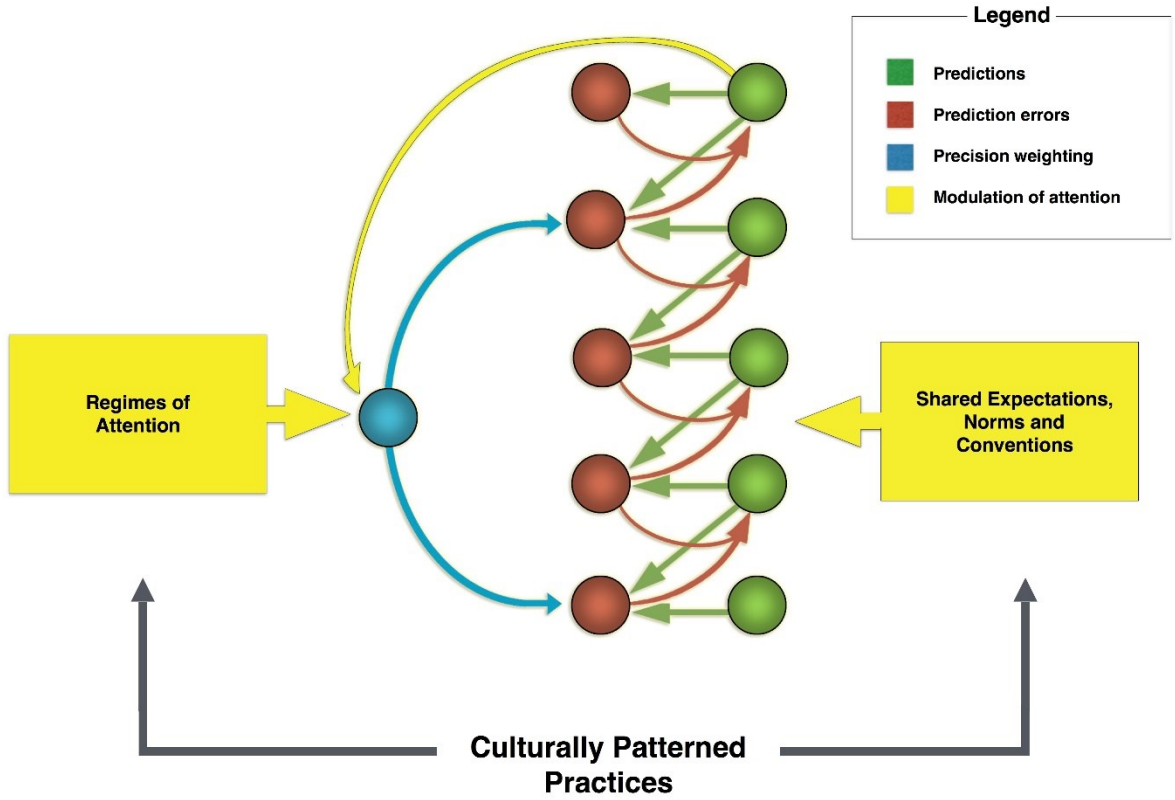


Fig. 6. *A diagram of the looping effects that mediate cultural affordance learning.* Regimes of attention, a central kind of patterned cultural practice, and higher level expectations encoded in higher levels of the cortical hierarchy, guide agents' attentional styles. In the free-energy framework, attention is modelled as precision-weighting and has the function of controlling activation across the various levels of the cortical hierarchy by

tuning the gain on error units (that is, they realize the function of gating effective connectivity in the brain). In turn, differences in how attention is deployed (through gating) lead to varying salience landscapes and to different expectations being encoded in the predictive hierarchy. Based in part on Figure 1 in Friston et al. (2014).

It has been argued that predictive processing models offer a plausible implementation for the neural-computational realization of affordance-responsiveness in the nervous system (Clark, 2015). As we shall see below, the free-energy model provides a mechanistic implementation of the dynamical gradient generation and consumption conception of affordance engagement examined above (Bruineberg & Rietveld, 2014). Free-energy is minimized through action and perception by the predictive processing hierarchy, which provides a mechanistic implementation of the descriptive-prescriptive aspect of affordances.

4. Cultural affordances and shared expectations

We lack comprehensive accounts of how the conventions that give rise to sociocultural affordances are successfully internalized, both as implicit knowing how and explicit knowing that. As Searle and others (Sterelny, 2007; Tomasello, 2014; Tuomela, 2007; but see Zahavi & Satne, 2015) have shown, and as our models suggests, it takes higher-order levels of intentionality, meta-communication, and perspective-taking in order for symbolic conventions to be used and manipulated—and for more complicated, self-referential thinking ("I know that she thinks that I believe that she intends to *X*," etc.), collective intentionality, and multiple orders of mindreading.

The question for the present essay is how this framework can be scaled up to account for cultural and social cognition and learning. The everyday phenomenology of affordances is one of possibilities for action and their variations; in other words, of *expecting* certain nested action possibilities and prescriptions for action. In effect, the phenomenology of affordances is a phenomenology of *expectations* about available and appropriate agent-environment couplings. The neural-computational models derived from the free-energy principle traffics in predictions and conditional probability distributions (called 'beliefs' in Bayesian probability theory, without any claim to correspond to the folk psychological notion). Arguably, the phenomenological correlate of these Bayesian beliefs can, at least at some (presumably higher) levels of the predictive hierarchy, be thought of as (or at least codetermine) *agent-level*

expectations. Our remarks below focus on clarifying how the social scaffolding of agents leads to their acquisition of representational content in regimes of shared attention.

4.1. Skilled intentionality and affordance competition

On the radical embodied view, the central feature of the dynamic relations between organisms and environment is the tendency of the organism to move towards an 'optimal grip' on the situation. The optima in question, as nearly everywhere in biology, are local optima, rather than a single global optimum. Under the free-energy framework, the 'optimal grip' can be understood as the pattern of action-perception that most minimizes variational free-energy. The free-energy minimizing dynamics of the predictive hierarchy might be described as a kind of weighted or biased competition between different affordances, the 'affordance competition' hypothesis (Cisek, 2007; Cisek & Kalaska, 2010; Pezzulo & Cisek, 2016). This model of action selection theorizes that the cognitive system appraises different trajectories for motor action simultaneously during action selection (that is, appraising a whole field of affordances in parallel and dynamically settling on the most salient affordance).

Sport science provides an illustration of this tendency toward optimal grip (Chow, Davids, Hristovski, Araújo, & Passos, 2011; Hristovski, Davids, Araújo, & Button, 2006; Hristovski, Davids, & Araujo, 2009). Studies of the dynamic interplay between a boxer's stance and position, and the action possibilities available to them as a function of stance and position, have shown that punching bags afford different kinds of strikes to boxers as a function of the distance between boxer and punching bag. Boxers tend to move their bodies to an optimal distance from the punching bag, specifically, one that affords the greatest variety of strikes. This is a case of moving towards optimal grip. When observing a painting, we also move our bodies and our gazes in a way that maximizes our grip on the scene or details observed. We might call such dynamic adaptive engagement with field of affordances in rolling cycles of action-perception 'skilled intentionality' (following Bruineberg & Rietveld, 2014; Merleau-Ponty, 2013/1945; Rietveld, 2008b, 2012).

Using the theoretical frameworks of dynamical systems and self-organization, Bruineberg and Rietveld (2014) have conceptualized this skilled intentionality as a kind of coping with the potentials that well up in the field of affordances, as a result of the dynamic relations between organism (with its phenotypical states, its states of action readiness, its concerns, etc.) and environment. More specifically, they suggest that skilled intentionality is the generation and reduction (or 'consumption') by the organism of a 'gradient' or potential tension in the field of affordances (which can be modeled using attractor dynamics). We

sketched this approach in sections 1.2. and 1.3., without the free-energy framework. The full significance of dissipative dynamics in the field of affordances can now be appreciated.

Affordances that are relevant to the organism at a given time (solicitations) drive system dynamics by soliciting rolling loops of action-perception and are prescribed and consumed or dissipated by those very dynamics (Tschacher & Haken, 2007). That is, solicitations are equivalent to potentials in the field of affordances, which act as attractors on the organism-environment dynamics, changing those affordances to which the organism is selectively open and receptive. The solicitations with which the organism engages, on this view, is the one that most effectively minimizes free-energy. Affect, attention, and affordances interact to sculpt a field of solicitations out of the total landscape of available affordances, adaptively and dynamically moving the organism towards an optimal grip on situations through action-perception. As the organism moves along a gradient toward an optimal grip, the gradient dissipates. The field of affordances thus changes dynamically along with perception-action and changes to states of the organism and environment. Responsiveness to the field, informed by states of the organism and environment, prescribe modes of optimal coupling. The radical embodied conception of cognition as skilled intentionality, then, can be modeled using systems theoretical models as a kind of selective responsiveness to salient available affordances or solicitations, modulated by states of the organism (concerns, interests, abilities) and states of the environment. This framework effectively bridges the descriptive levels of phenomenology, system dynamics, and cognitive functions or mechanisms.

To date, most work on affordances has focused on motor control and basic behaviors related to dynamical embodied coping (e.g., Chemero, 2009; Cisek & Kalaska, 2010; Pezzulo & Cisek, 2016). For a theory of cultural affordances, the notion of affordances must be extended to more complex features of the social and cultural niche inhabited by humans (Bruineberg & Rietveld, 2014; Heft, 2001; Rietveld & Kiverstein, 2014). Quintessential human abilities like language, shared intentionality, and mind-reading/perspective-taking emerge from human forms of life and are patterned by human sociocultural practices (Roepstorff et al., 2010), which in turn involve sophisticated forms of social cognition. We live in a landscape of cultural affordances.

4.2. Shared expectations, local ontologies, and cultural affordances

The upshot of our discussion so far is a general concept of skilled intentionality as selective engagement with a field of affordances supported by embodied generative models. Skilled intentionality is a graded phenomenon. At one extreme, skilled intentionality consists in

contentless direct coping. It has been suggested that this most basic form of intentionality, which Hutto and Satne (2015) call 'ur-intentionality', acquires its tendencies for selective targeted engagement with the world in a 'teleosemiotic' process shaped by evolutionary history.[6] At this extreme, the only information (and affordances) needed are of the natural kind (exploitable reliable correlation). At the other extreme, we find stereotypical human intentionality, that is, symbolically dense and strongly content-involving forms of collectively and conventionally rooted intentionality (Kiverstein & Rietveld, 2015), which involves conventional information and affordances. This is a spectrum, and all points between these extremes are viable (at least prima facie). The teleological basis of this variation might be the needs, concerns, and abilities relevant to a given form of life, (Kiverstein & Rietveld, 2015; Rietveld & Kiverstein, 2014), in specific social niches with their own idiosyncratic shared representations, symbols, etc.

Our claim here is that cultural affordances (especially conventional ones) form a coordinated affordance landscape, which is enabled by sets of embodied expectations that are shared by a given community or culture. Social niches and cultural practices generally involve not isolated, individual affordances or expectations but local landscapes that give rise to and depend on shared expectations. We submit that these shared expectations—implemented in the predictive hierarchies, embodied in material culture, and enacted in patterned practices—contribute to the constitution of the landscape of affordances that characterizes a given community or culture. Indeed, shared expectations modulate the specific kinds of intentionality that are effective in a given community, determining the forms taken by skilled intentionality, especially the shared skilled intentionality of the kind that constitutes a patterned sociocultural practice.

Patterned practices are specific ways of doing joint activities in domain-specific material-discursive environments (Roepstorff et al., 2010). Echoing recent work on the natural origin of semantic content (Hutto & Satne, 2015; Sterelny, 2015), we hypothesize that such ontologies, as socially shared and embodied expectations, come to be acquired by the

---

[6] 'Teleosemiotics' is teleosemantics minus the semantics, that is, using the teleosemantic framework developed by Millikan (1984, 2004, 2005) to explain how organisms develop selective intentional response tendencies without trying to provide thereby an account of semantic content (Hutto & Myin, 2013). See also Kiverstein & Rietveld (2015) for a complementary account of minimal intentionality as a contentless form of skilled intentionality.

individual agent through their participative immersion in specific patterned practices available in multi-agent, symbolically and linguistically mediated forms of social life.

Building on work in cognitive science as well as by Hacking (1995; Hacking, 1999; Hacking, 2002; Hacking, 2004), Kirmayer and colleagues have argued for an embodied, enactivist approach to the study of the multilevel feedback or 'looping' effects involved in jointly-mediated narratives, metaphors, forms of embodiment, and mechanisms of attention (Kirmayer, 2008; Kirmayer, 2015; Kirmayer & Bhugra, 2009; Kirmayer & Gold, 2012; Seligman & Kirmayer, 2008). In human life, the regularities to which agents are sensitive are densely mediated (and often constituted) by cultural symbols, narratives, and metaphors, which may explicitly reference or tacitly assume particular ontologies. These mechanisms shape social experience and in turn are shaped by broader social contexts.

Elsewhere, we have suggested that local, culturally specific ontologies can be understood as sets of shared expectations (Kirmayer & Ramstead, 2017). A 'local ontology' can be defined as a mode of collective expectation: agents expect the sociocultural world to be disclosed in certain ways rather than others, and to afford certain forms of action-perception and nested variations to the exclusion of others. A local ontology, then, is a set of expectations that are shared by members of a cultural community. We claim that these sets of shared expectations are installed in agents through patterned practices that result in enculturation and enskillment. In the framework explored above, these ontologies codetermine the exact affordances that are available in a given niche, for they prescribe specific ways of being, thinking, perceiving, and acting in context that are situationally appropriate.

These local ontologies need not be explicitly formulated as metaphysical theories. They are more often implicit and acquired through participation in patterned practices and the enactment of customs and rituals, or embodied in the social material reality itself (as symbols, places, stories). Such distinctively human practices take place in social niches rich with narratives, symbols, and customs, which enable individuals to respond cooperatively and, at times, to infer other agents' states of mind. Such practices may underlie everyday processes of person-perception. For example, as noted in the introduction, by age 5, children have acquired local ontologies and categories of personhood—which reproduce the dominant set of biases, expectations, and representations of their cultures—showing preference for dominant group culture often without being explicitly taught to do so, and despite their caregivers not consciously holding such views, even when these biases are not consonant with their minority identities (Clark & Clark, 1939; Kinzler & Spelke, 2011). These tacit views of others may arise both from the ways in which local niches are structured by social norms and conventions and

from regimes of attention and interpersonal interactions shaped by cultural practices (Richeson & Sommers, 2016). Biases in person-perception will, in turn, influence subsequent social interaction and cooperative niche construction in a cognitive-social loop (Sacheli et al., 2015).

As discussed above, a number of theorists of embodied cognition have criticized the view that intersubjective interactions require that human beings be endowed with the capacity for mind-reading, opting instead for an explanation in terms of embodied practices and coupling (Fuchs & De Jaegher, 2009; Gallagher, 2001, 2008). Although we readily grant the importance of such embodied coping for basic minds on which more elaborate cognition can be scaffolded, we advocate a middle ground that posits both embodied contentless abilities and more contentful mindreading abilities (Michael et al., 2014; Sterelny, 2015; Tomasello, 2014; Veissière, 2016). Indeed, the framework we have proposed, which posits predictive processing hierarchies apt to engage with both natural and conventional information and affordances, can accommodate both modes of cognition. The view that human societies rely on explicit and implicit forms of mindreading does not commit us to intellectualism or to a strong content-involving view. The shared enactment of meaning, involving expectations about other agents, comes to constitute the shared, taken-for-granted meaning of local worlds, which in turn feeds back, in a kind of looping effect, to developmentally ground and scaffold the enactments of meaning by individual agents, by altering the shared expectations that are embodied and enacted in the social niche (Kirmayer, 2015). These shared ontologies shape experience by changing the abilities and styles of action-perception of encultured agents.

4.3. Shared expectations and implicit learning

We have already appealed to Grice's theory of meaning to clarify some aspects of affordances. Affordances come in a spectrum, ranging from those that depend only on reliable correlation to those that depend on shared sets of expectations. Grice's account, as improved by others (Levinson, 2000; Sperber & Wilson, 1986; Tomasello, 2014), can help account for how we successfully learn to detect and selectively respond to context in situations that involve higher order contextual appraisal, including perspective-taking and reading of other's goal-directed intent and actions. In higher-order, rule-governed semiotic contexts, the actual presence of others is not necessary for inferences to be made about the 'correctness' of affordances in terms of their correspondence to others' expectations, norms or conventions. The general internalized idea of how others would interpret a situation and context (or how a culturally competent actor would respond) suffices for 'meaning' to be derived or inferred.

Most of us have never been explicitly taught precisely how to behave, sit, move, speak, take turns, and interact with others in shared spaces such as metros, elevators, hallways, airplanes, university classrooms, bars, dance floors, janitors' closets, or the many other spaces we know not to enter. As mentioned in the introduction to this essay, children acquire the dominant social norms and appropriate behavioral repertories and responses without explicit instruction. Although we do occasionally receive explicit instructions, these do not seem necessary for normal social functioning; as Varela (1999) pointed out, we have acquired the implicit 'know how' to act appropriately. That is, human beings acquire characteristic, stereotypical ways of doing and being in response to social contexts; in a sense, each of these constitutes habitual 'micro-selves' as we variously engage the world as our 'getting-on-the-bus-self' to our 'having-lunch-self,' etc., where each self is a style of situationally adequate and socially appropriate coupling to a context. How do we acquire the ability to selectively detect and respond to such sociocultural affordances? Or to rephrase the question in anthropological terms: How do we come to be socialized or enculturated for participation in shared worlds of expectations?

The highly stable conformity of behavior in all of these contexts goes beyond direct imitation (Michael et al., 2014). Many everyday situations involve coordinated action among many participants. Although some forms of coordinated group action can occur entirely through individual responses to local impersonal affordances (e.g. the swarming of birds), in order to read and master the social cues and scripts in complex human settings, the actors involved need to grasp the situation from the perspective of other actors. This perspective-taking is essential if each actor's appraisal of the situation is to have any counterfactual depth with regard to explicit social norms (e.g., inferring that one's behaving differently would fail to conform to others' *expectations* about correct behavior). However, as argued above, in some instances this perspective-taking might not involve explicit, content-involving processes; the expectations might simply be encoded and leveraged for the generation of adaptive behavior without mentalistic assumptions being made about agents at an explicit, conscious level. Thus, in any case, for a given space to afford the same engagements to a given population, that community must come to share a set of collective expectations—indeed, shared expectations about *others' expectations about our expectations*, and so forth.

5. Regimes of shared attention and shared intentionality

The framework we have outlined for cultural affordances allows us to reconsider the natural origins of content. We hypothesize that the central mechanism whereby cultural affordances

are acquired, especially conventional, content-involving affordances, consists in the looping or feedback relations between shared intentionality and shared attention. Shared intentionality is enacted in various concrete, materially embedded cultural practices and embodied as shared sets of expectation. Shared attention is one such form of shared intentionality. We suggest that shared attention is crucial because directed attention modulates the agent's selective engagement with the field of affordances. Given the nature of the predictive hierarchy, to wit, to extract explicit and implicit statistical information, directing an agent's attention is tantamount to determining which expectations (Bayesian prior beliefs) will be encoded in the hierarchy. This, in turn, leads to different sets of abilities being implemented by the gating mechanisms of the predictive hierarchy. Under the free-energy principle, action-perception is guided attention (precision-weighting), and the gating process that is realized by attention itself rests on the expectations encoded in the generative models embodied by the organism. These high-level expectations about precision, which modulate allocations of attention (and thereby determine action-perception through gating) are leveraged to guide skillful intentional behavior. The sets of expectations embodied and enacted by organisms change the field of affordances. This mechanism, we submit, is exploited by culture in the acquisition of cultural affordances.

5.1. Gating, abilities, and affordances

In the framework outlined above, we followed Rietveld and Kiverstein (2014) in defining an affordance as a relation between a set of features or aspects of the organism's material environment and the abilities available in that organism's form of life. We are now in a position to better define an *ability* in terms of a gating control pattern, that is, a sequenced or coordinated process. An ability is simply the capability of an organism to coordinate its action-perception loops to skillfully engage an affordance in a way that is optimal under the free-energy principle. An ability, then, in the free-energy framework, includes a pattern of attention, in the specific sense employed by the free-energy framework. We use the term 'attention' not in the folk-psychological sense, as that effort or mechanism that allows us to attend to specific aspects of experience, but as the mechanism of precision-weighting that mediates neural gating and allows the agent to engage with specific affordances in action-perception cycles. Attention, in our technical sense, therefore modulates effective connectivity and, as such, determines the trajectories taken by the rolling cycles of action-perception. Typically, in the case of human agents, such patterns of attention are acquired over development.

We conjecture that we acquire our distinctively human abilities from our dense histories of temporally coordinated social interaction and shared cultural practices (Roepstorff, 2013; Tomasello, Carpenter, Call, Behne, & Moll, 2005). Attentional processes are central to this enculturation and installation of shared semantic content. In particular, the landscape of affordances available to the infant is sculpted, through joint-attentional practices that reflect sociocultural norms, into a *field* of relevant solicitations. Thus, participation in patterned practices allows the installation of socially, culturally, and situationally specific expectations, which, once acquired, determine agent allocations of attention (the acquisition of abilities) and, as a result, guide action-perception.

Joint (and, eventually, shared) attentional processes (Tomasello, 2014) provide a central mechanism through which the individual is molded to conform to specific group expectations and participate in forms of cooperative action. Joint and shared attention alter the field of affordances by directing the agent to engage with specific affordances, marking them out as relevant, and making them more salient. Given the nature of the predictive hierarchy, that is, to automatically extract statistical information about the distal world in its dynamic engagement (in action-perception), the agent will encode the regularities of the solicitations that it engages (that is, the relevant affordances to which it is directed in joint and shared attention). Of course, local practices of joint and shared attention themselves depend on agents sharing sets of expectations—the same expectations that become encoded by agents as they participate in these practices. Through participation in patterned cultural practices that direct attention in specific ways, the agent acquires sets of expectations that gave rise, in the first instance, to (earlier versions of) that very form of cooperative action (see Figure 6). Cultural affordances are thus mediated by recursive regimes of shared attention, of which joint-attention is a special, signal case (Tomasello, 2014).

The study of everyday social interactions reveals how regimes of joint attention shape our understanding and sensory experiences of being in our worlds. For example, Goffman, who pioneered studies of face-to-face interaction in modern societies, showed how the ‘anonymized’, ‘surface character’ of life in cities is routinized through what he called ‘civic inattention’—that is, through the many ways in which strangers avert their gazes, avoid conversations or physical contact, and reinforce private boundaries in the public sphere (Goffman, 1971, p. 385). We can follow Goffman’s lead to consider how different regimes of shared and joint-attention mediate lived experiences of meaning and being. Civic inattention, for example, is a specific regime of attention, but it is certainly not an absence of attention. In Goffman’s ‘Invisible City’ model, attentional resources are mobilized to *not* pay attention to

certain features of the world, particularly other agents caught in a symbolically-marked game of allegiances that renders them strange or invisible.

5.2. Looping the loop: Regimes of shared attention and skilled intentionality

As we have seen above, in the predictive processing scheme, attention, understood as precision-weighting of prediction error signals, is a central mechanism behind the dynamical trajectory of action-perception. The expectations about precision that guide action-perception are acquired in ontogeny and stored as high-level priors, which have the effect of arbitrating the balancing act between top-down prediction and bottom-up error signals. It follows that one pathway by which cultural affordances may be transmitted is through the manipulation of attention. This may occur in a variety of ways including what we might call '*regimes of shared attention*'. In the model of affordances outlined above, this kind of attentional modulation involves carving a local field of affordances out of the larger landscape of available affordances through social practices. Local environments and their associated practices are designed to solicit particular patterns of coordinated attention from participants (Clark, 2015; Kirchhoff, 2015a). In effect, these patterns act as dynamical attractors on the field of affordances, directing action-perception in some ways rather than others (Juarrero, 1999).

In this light, one can view social norms and conventions as devices to reduce mutual uncertainty, that is, consonantly with the free-energy framework, as entropy-minimizing devices (Colombo, 2014). One must know 'what is in the minds' of others (such as what one would see and how one would interpret another's action generally and in context) in order to make a successful inference (both explicit, content-involving or implicit, correlational inferences) about other agents in each situation. Goffman (1971) was hinting a similar processes with his comments on the 'faces' we learn to perform when we interact with others in different situations. We can be a mentor in one situation, and a mentee in another; a father in one and a friend in another. In Goffman's famous comments on interaction in public, he describes (using other terms), how certain spaces afford more 'backstage', 'off-screen' performances than others. The privacy of the home affords such relaxed 'off-stageness', and the bedroom and bathroom even more so. All these instances require inferential mindreading or perspective-taking, that is, inferences about the presence or absence of other agents and their expectations as a normative guide for how one can behave. None of this depends specifically of whether these inferences consist in explicit mindreading or more implicit forms of embodied coupled enactments—both are compatible with our framework.

Now, we might suppose that the distinctly human abilities with which we are endowed result simply from better evolved predictive machinery, that is, more computationally powerful predictive hierarchies (Conway & Christiansen, 2001). However, as we argued above, in human ontogeny, it is more likely that affordances are learned through regimes of imitation, repetition, positive and negative conditioning, and culturally selective forms of attention (Banaji & Gelman, 2013; Meltzoff & Prinz, 2002; Roepstorff & Frith, 2004; Veissière, 2016; Whitehouse, 2002; Whitehouse, 2004). The capacity for cultural learning may itself be a cultural innovation (Heyes, 2012). Indeed, the feedback or looping mechanisms between cultural practices of scaffolding individual attention (what we called regimes of attention) are themselves determined by the local ontologies (shared sets of expectations) and abilities (acquired patterns of attention and gating) of agents in that community. Repetition and reiteration of patterns of social and technological interaction, as well as rewards for 'correct' inferences that denote an adequate grasp of relevance, prescription, and proscription (e.g., when a child 'gets' that some *X* means some *Y*, or figures out an 'appropriate' combination of meaningful elements in any given context), come to shape attentional mechanisms in ontogeny, and assist the child in successfully inferring a set of rules and categories (the culturally sanctioned sets of shared expectations).

Joint attention is usually understood as occurring in a dyad of two people, or between agents in direct interactional spheres of communication, gaze-following, finger-pointing, or other verbal or non-verbal cues (Tomasello, 2014; Vygotsky, 1978). To address more complex social situations, it is useful to revise current sociocognitive models of joint-attention to encompass fundamentally triadic situations in which 'the third' is the socially constituted niche of affordances, supported by local ontologies and abilities.

Shared human intentionality is sufficient to project joint attention to larger groups in the process of forming joint goals and inferring from joint expectations. Crucially, it commonly takes place without any direct interaction from members, in the many routinized, anonymous, symbolically and linguistically mediated forms of sociality, including engagement with social institutions.

To go beyond the 'toy models' of dyadic joint attention to grasp the process of culture transmission we need to study the dynamics of 'designer environments' (Goldstone, Landy, & Brunel, 2011; Salge, Glackin, & Polani, 2014). Human beings pattern their environments in a process of recursive niche construction, which in turn modulates the attributions of attention in individual agents, leading them to acquire certain sets of priors rather than others, in what Sterelny (2003) has called 'incremental downstream epistemic engineering'. This incremental

process of constructing our own collective, epistemic niches, involves a kind of bootstrapping in which symbolically and linguistically mediated forms of human communication can be modeled as forms of re-entrant processing. Linguistically abled human beings produce patterned, structured outputs that become part of the material environment, and are subsequently picked up and further processed by other agents in ways that stabilize and elaborate a local social world (Clark, 2006, 2008). Indeed, human-constructed environments, which shape agent expectations and guide patterns of attention, can be viewed as another level of the generative statistical model of the niche, which human beings leverage to guide intelligent behavior in their sociocultural symbolically- and linguistically-laden niches (Clark, 2015; Kirchhoff, 2015a). The prior knowledge that is leveraged in action-perception is thus encoded in multiple level and sites: in the hierarchical neural networks, in the organism's phenotype (over phylogeny and ontogeny), and in patterned sociocultural practices and designer environments.

Thus, our suggestion is that regimes of attention, which mediate the acquisition of cultural affordances (both natural and conventional), are enacted through patterned practices (especially those which modulate the allocation of attention) and are embodied in sundry ways: in the predictive hierarchies of individual agents in a community, as encoded sets of expectations, and in the concrete social and cultural world, as constructed human environments, designed to solicit certain expectations and direct attention.

6. Conclusion

We have outlined a framework for the study of cultural affordances in terms of neural models of predictive processing and social practices of niche construction. This approach can help account for the multilevel forms of affordance learning and transmission of affordances in socially and culturally shared regimes of joint-attention and clarify one of the central mechanisms that can explain the natural origins of semantic content. The concepts of affordance and skilled intentionality in ecological, radical embodied, and enactivist cognitive science can be supplemented with an account of the nature of affordances in the humanly constructed sociocultural niches. Turning to cultural niche construction, we argued in favor of a conception of local ontologies as sets of shared expectations acquired through the immersive engagement of the agent in feedback looping relations between shared intentionality (in the form of shared embodied expectations) and shared attention (modulated by regimes of attention). We elaborated Grice's account of meaning by highlighting the dependence of selective responsiveness to cultural affordances on shared and joint intentionality, modes of

conventionality and social normativity. We ended with an account of the patterned regimes of attention and modes of social learning that might lead to the acquisition and installation of such ontologies and affordances, leading to agent enculturation and enskillment. We hope that our proposal of a framework for the study of cultural affordances will spur further research on multilevel, recursive, nested affordances and the expectations on which they depend.

Box 1. *Basic concepts of a framework for cultural affordances*

Affordance: A relation between a feature or aspect of organisms' material environment and an ability available in their form of life. (Bruineberg & Rietveld, 2014; Chemero, 2003, 2009; Rietveld & Kiverstein, 2014)

Landscape of affordances: The total ensemble of available affordances for a population in a given environment. This landscape corresponds to what evolutionary theorists in biology and anthropology call a 'niche'. (Bruineberg & Rietveld, 2014; Rietveld, 2008a, 2008c; Rietveld et al., 2013; Rietveld & Kiverstein, 2014)

Field of affordances: Those affordances in the landscape with which the organism, as an autonomous individual agent, dynamically copes and intelligently adapts. The field refers to those affordances that actually engage the individual organism because they are salient at a given time, as a function of the interests, concerns, and states of the organism. (Bruineberg & Rietveld, 2014; Rietveld, 2008a, 2008c; Rietveld & Kiverstein, 2014)

Cultural affordance: The kind of affordance that humans encounter in the niches that they constitute. There are two kinds of cultural affordances: natural and conventional affordances.

Natural affordance: Possibilities for action (i.e. affordances), the engagement with which depends on the exploitation or leveraging by an organism of 'natural information', that is, reliable correlations in its environment, using its set of phenotypical and encultured abilities (roughly what Grice meant by 'natural meaning'). (Piccinini, 2015; Piccinini & Scarantino, 2011)

Conventional affordance: Possibilities for action, the engagement with which depends on agents' skillfully leveraging explicit or implicit expectations, norms, conventions, and cooperative social practices in their ability to correctly infer (implicitly or explicitly) the culturally specific sets of expectations of which they are immersed. These are expectations about how to interpret other agents, and the symbolically and linguistically mediated social world. (Satne, 2015; Scarantino, 2015; Scarantino & Piccinini, 2010; Tomasello, 2014)

References

Aboud, F. E., & Amato, M. (2008). Developmental and socialization influences on intergroup bias. In S. M. Quintana & C. McKown (Eds.), *Handbook of race, racism, and the developing child* (pp. 65-85). Hoboken, NJ: John Wiley & Sons.

Adams, R. A., Bauer, M., Pinotsis, D., & Friston, K. J. (2016). Dynamic causal modelling of eye movements during pursuit: Confirming precision-encoding in V1 using MEG. *Neuroimage, 132*, 175-189.

Anderson, M. L. (2014). *After phrenology : neural reuse and the interactive brain*. Cambridge, MA; London, England: The MIT Press.

Anderson, M. L., & Chemero, T. (2013). The problem with brain GUTs: conflation of different senses of "prediction" threatens metaphysical disaster. *Behavioral and Brain Sciences, 36*(3), 204-205.

Badcock, P. B. (2012). Evolutionary systems theory: a unifying meta-theory of psychological science. *Review of General Psychology, 16*(1), 10-23.

Banaji, M. R., & Gelman, S. A. (2013). *Navigating the social world: What infants, children, and other species can teach us*. Oxford: Oxford University Press.

Bar, M. (2011). *Predictions in the brain: Using our past to generate a future*. Oxford: Oxford University Press.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol., 59*, 617-645.

Bechtel, W. (2007). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*: Psychology Press.

Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in human neuroscience, 8*. doi:doi.org/10.3389/fnhum.2014.00599

Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology, 15*(2), 181-195.

Chemero, A. (2009). Radical embodied cognition. In: Cambridge, MA: MIT Press.

Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. *Synthese, 191*(2), 127-153.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT press.

Chow, J. Y., Davids, K., Hristovski, R., Araújo, D., & Passos, P. (2011). Nonlinear pedagogy: Learning design for self-organizing neurobiological systems. *New Ideas in Psychology, 29*(2), 189-200.

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical transactions of the Royal Society B: Biological sciences, 362*(1485), 1585-1599.

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual review of neuroscience, 33*, 269-298.

Claidière, N., Scott-Phillips, T. C., & Sperber, D. (2014). How Darwinian is cultural evolution? *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 369*(1642), 20130368.

Clark, A. (1998). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT press.

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences, 10*(8), 370-374.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. New York: Oxford University Press

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181-204.

Clark, A. (2015). *Surfing uncertainty: prediction, action, and the embodied mind*. Oxford: Oxford University Press.

Clark, K. B. (1963). *Prejudice and your child*. Boston: Beacon Press.

Clark, K. B., & Clark, M. K. (1939). The development of consciousness of self and the emergence of racial identification in Negro preschool children. *The Journal of Social Psychology, 10*(4), 591-599.

Colombo, M. (2014). Explaining social norm compliance. A plea for neural representations. *Phenomenology and the Cognitive Sciences, 13*(2), 217-238.

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences, 5*(12), 539-546.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*: Oxford University Press.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience, 8*(12), 1704.

Dawson, M. R. (2013). *Mind, body, world: Foundations of cognitive science*. Edmonton: Athabasca University Press.

De Haan, S., Rietveld, E., Stokhof, M., & Denys, D. (2013). The phenomenology of deep brain stimulation-induced changes in OCD: An enactive affordance-based model. *Frontiers in human neuroscience, 7*(653), 1-14.

den Ouden, H. E., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *Journal of Neuroscience, 30*(9), 3210-3219.

Di Paolo, E. (2009). Extended life. *Topoi, 28*(1), 9.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences, 4*(4), 429-452.

Di Paolo, E. A., & Thompson, E. (2014). The enactive approach. In L. E. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 68-78). London: Routledge.

Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.

Eliasmith, C. (2005). A new perspective on representational problems. *Journal of Cognitive Science, 6*(97), 123.

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in human neuroscience, 4*, 215.

Fodor, J. A. (1975). *The language of thought* (Vol. 5): Harvard University Press.

Frijda, N. H. (1986). *The emotions*. Cambridge: Cambridge University Press.

Frijda, N. H. (2007). *The laws of emotions*. Mahwah, NJ: Erlbaum.

Friston, K. (2012a). A free energy principle for biological systems. *Entropy, 14*(11), 2100-2121.

Friston, K. (2012b). Predictive coding, precision and synchrony. *Cognitive neuroscience, 3*(3-4), 238-239.

Friston, K. (2013a). Active inference and free energy. *Behavioral and Brain Sciences, 36*(3), 212-213.

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

Friston, K., & Frith, C. (2015a). A duet for one. *Consciousness and cognition, 36*, 390-405.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical transactions of the Royal Society B: Biological sciences, 364*(1521), 1211-1221.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Human brain mapping, 2*(1-2), 56-78.

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127-138.

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain connectivity, 1*(1), 13-36.

Friston, K. J., Bastos, A. M., Pinotsis, D., & Litvak, V. (2015). LFP and oscillations—what do they tell us? *Current opinion in neurobiology, 31*, 1-6.

Friston, K. J., & Frith, C. D. (2015b). Active inference, communication and hermeneutics. *cortex, 68*, 129-143.

Friston, K. J., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris, 100*(1), 70-87.

Friston, K. J., Levin, M., Sengupta, B., & Pezzulo, G. (2015). Knowing one's place: a free-energy approach to pattern regulation. *Journal of the Royal Society Interface, 12*(105).

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., . . . Bestmann, S. (2012). Dopamine, affordance and active inference. *PLOS Computational Biology, 8*(1), e1002327.

Friston, K. J., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. *The Lancet Psychiatry, 1*(2), 148-158.

Frith, C. (2007). *Making up the mind: How the brain creates our mental world*. Malden, MA: Blackwell John Wiley & Sons.

Froese, T., & Di Paolo, E. A. (2011). The enactive approach: Theoretical sketches from cell to society. *Pragmatics & Cognition, 19*(1), 1-36.

Fuchs, T., & De Jaegher, H. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. . *Phenomenology and the Cognitive Sciences, 8*(4), 465-486.

Fuentes, A. (2014). Human evolution, niche complexity, and the emergence of a distinctively human imagination. *Time and mind, 7*(3), 241-257.

Gallagher, S. (2001). The practice of mind. Theory, simulation or primary interaction? *Journal of Consciousness Studies, 8*(5-6), 83-108.

Gallagher, S. (2008). Inference or interaction: social cognition without precursors. *Philosophical Explorations, 11*(3), 163-174.

Gibson, J. J. (1979). *The ecological approach to visual perception*: Psychology Press.

Goffman, E. (1971). *Relations in Public: Microstudies of the Public Order*. New York, NY: Basic Books.

Goldstone, R. L., Landy, D., & Brunel, L. C. (2011). Improving perception to make distant connections closer. *Frontiers in Psychology, 2*(385). doi:doi:10.3389/fpsyg.2011.00385

Grice, H. P. (1957). Meaning. *The philosophical review, 66*(3), 377-388.

Grice, H. P. (1969). Utterer's meaning and intention. *The philosophical review, 78(2)*(2), 147-177.

Grice, H. P. (1971). Intention and uncertainty. *Oxford, Oxford University Press*.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Grush, R. (2001). The semantic challenge to computational neuroscience. In P. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and method in the neurosciences* (pp. 155-172). Pittsburgh: University of Pittsburgh Press.

Hacking, I. (1995). The looping effect of human kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal Cognition: A Multidisciplinary Debate* (pp. 351-383). Oxford: Oxford University Press.

Hacking, I. (1999). *The social construction of what?* : Harvard university press.

Hacking, I. (2002). *Historical Ontology*. Cambridge, MA: Harvard University Press.

Hacking, I. (2004). Between Michel Foucault and Erving Goffman: between discourse in the abstract and face-to-face interaction. *Economy and society, 33*(3), 277-302.

Haugeland, J. (1990). The intentionality all-stars. *Philosophical perspectives, 4*, 383-427.

Heft, H. (2001). *Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James's radical empiricism*: Psychology Press.

Heidegger, M. (2010/1927). *Being and time*: SUNY Press.

Heyes, C. (2012). New thinking: the evolution of human cognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 367*(1599), 2091-2096.

Hirschfeld, L. A. (1998). *Race in the making: Cognition, culture, and the child's construction of human kinds*: MIT Press.

Hohwy, J. (2014). *The predictive mind*. Oxford: Oxford University Press.

Howes, D. (2011). Reply to Tim Ingold. *Social Anthropology, 19*(3), 318-322.

Hristovski, R., Davids, K., Araújo, D., & Button, C. (2006). How boxers decide to punch a target: emergent behaviour in nonlinear dynamical movement systems. *Journal of sports science & medicine, 5*(CSSI), 60.

Hristovski, R., Davids, K. W., & Araujo, D. (2009). Information for regulating action in sport: metastability and emergence of tactical solutions under ecological constraints.

In *Perspectives on cognition and action in sport* (pp. 43-57): Nova Science Publishers, Inc.

Huneman, P., & Machery, E. (2015). Evolutionary psychology: issues, results, debates. In *Handbook of Evolutionary Thinking in the Sciences* (pp. 647-657): Springer.

Hutchins, E. (2014). The cultural ecosystem of human cognition. *Philosophical Psychology, 27*(1), 34-49.

Hutto, D., & Satne, G. (2015). The natural origins of content. *Philosophia, 43*(3), 521-536.

Hutto, D. D., Kirchhoff, M. D., & Myin, E. (2014). Extensive enactivism: why keep it all in? *Frontiers in human neuroscience, 8*, 706.

Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. Cambridge, MA: MIT Press.

Ingold, T. (2000). *The perception of the environment essays on livelihood, dwelling and skill*. New York: Routledge.

Ingold, T. (2001). From the transmission of representations to the education of attention. In H. Whitehouse (Ed.), *The debated mind: Evolutionary psychology versus ethnography* (pp. 113-153). Oxford: Berg Publishers.

Juarrero, A. (1999). *Dynamics in action*. Cambridge, MA: MIT Press.

Kelly, D., Faucher, L., & Machery, E. (2010). Getting rid of racism: Assessing three proposals in light of psychological evidence. *Journal of Social Philosophy, 41*(3), 293-322.

Kinzler, K. D., & Spelke, E. S. (2011). Do infants show social preferences for people differing in race? *Cognition, 119*(1), 1-9.

Kirchhoff, M. (2015a). Experiential fantasies, prediction, and enactive minds. *Journal of Consciousness Studies, 22*(3-4), 68-92.

Kirchhoff, M. (2015b). Species of realization and the free energy principle. *Australasian Journal of Philosophy, 93*(4), 706-723.

Kirchhoff, M. (2016). Autopoiesis, free energy, and the life–mind continuity thesis. *Synthese*, 1-22. doi:doi:10.1007/s11229-016-1100-6

Kirmayer, L., & Ramstead, M. (2017). Embodiment and Enactment in Cultural Psychiatry. In *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*: MIT Press

Kirmayer, L. J. (2008). Culture and the metaphoric mediation of pain. *Transcultural Psychiatry, 45*(2), 318-338.

Kirmayer, L. J. (2015). Re-visioning psychiatry: Toward an ecology of mind in health and illness. In L. J. Kirmayer, R. Lemelson, & C. Cummings (Eds.), *Re-visioning psychiatry: cultural phenomenology, critical neuroscience and global mental health* (pp. 622-660). New York: Cambridge University Press.

Kirmayer, L. J., & Bhugra, D. (2009). Culture and mental illness: social context and explanatory models. *Psychiatric diagnosis: patterns and prospects. New York: John Wiley & Sons*, 29-37.

Kirmayer, L. J., & Gold, I. (2012). Re-socializing psychiatry. *Critical neuroscience: A handbook of the social and cultural contexts of neuroscience*.

Kiverstein, J., & Rietveld, E. (2013). Dealing with context through action-oriented predictive processing. *Frontiers in Psychology, 3*, 421.

Kiverstein, J., & Rietveld, E. (2015). The primacy of skilled intentionality: on Hutto & Satne's the natural origins of content. *Philosophia, 43*(3), 701-721.

Kok, P., Brouwer, G. J., van Gerven, M. A., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience, 33*(41), 16275-16284.

Kok, P., Jehee, J. F., & De Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron, 75*(2), 265-270.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT press.

Machery, E., & Faucher, L. (2005). Why do we think racially? In *Handbook of categorization in cognitive science* (pp. 1009-1033): Elsevier.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information, henry holt and co*. San Francisco: W. H. Freeman.

Mead, M. (1975). *Growing up in New Guinea: A comparative study of primitive education*. New York: Morrow.

Meltzoff, A. N., & Prinz, W. (Eds.). (2002). *The imitative mind: Development, evolution and brain bases*. Cambridge: Cambridge University Press.

Merleau-Ponty, M. (1968/1964). *The visible and the invisible* (A. Lingus, Trans.). Evanston, IL: Northwestern University Press.

Merleau-Ponty, M. (2013/1945). *Phenomenology of perception*: Routledge.

Michael, J., Christensen, W., & Overgaard, S. (2014). Mindreading as social expertise. *Synthese, 191*(5), 817-840.

Miłkowski, M. (2013). *Explaining the computational mind*. Cambridge, MA: MIT Press.

Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*: MIT press.

Millikan, R. G. (2004). *Varieties of meaning*. Cambridge, MA: MIT press.

Millikan, R. G. (2005). *Language: A biological model*: Oxford University Press on Demand.

Moore, R. (2013). Social learning and teaching in chimpanzees. *Biology & Philosophy, 28*(6), 879-901.

Noë, A. (2004). *Action in perception*: MIT press.

O'Brien, G., & Opie, J. (2004). Notes toward a structuralist theory of mental representation. In *Representation in mind* (pp. 1-20): Elsevier.

O'Brien, G., & Opie, J. (2004). Notes toward a structuralist theory of mental representation. In H. Clapin, P. Staines, & P. Slezak (Eds.), *Representation in mind: New approaches to mental representation* (pp. 1-20).

O'Brien, G., & Opie, J. (2009). The role of representation in computation. *Cognitive processing, 10*(1), 53-62.

O'Brien, G., & Opie, J. (2015). Intentionality lite or analog content? *Philosophia, 43*(3), 723-729.

Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). *Niche construction: The neglected process in evolution*. Princeton: Princeton University Press.

Park, H.-J., & Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science, 342*(6158).

Pauker, K., Williams, A., & Steele, J. (2016). Children's racial categorization in context. *Child development perspectives, 10*(1), 33-38.

Petitot, J., Varela, F. J., Pachoud, B., & Roy, J.-M. (Eds.). (1999). *Naturalizing phenomenology: Issues in contemporary phenomenology and cognitive science*: Stanford University Press.

Pezzulo, G., & 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.

Piccinini, G. (2015). *Physical computation: A mechanistic account*: OUP Oxford.

Piccinini, G., & Scarantino, A. (2011). Information processing, computation, and cognition. *Journal of biological physics, 37*(1), 1-38.

Putnam, H. (1975). *Mind, language, and reality*. Cambridge: Cambridge University Press.

Richeson, J. A., & Sommers, S. R. (2016). Toward a social psychology of race and race relations for the twenty-first century. *Annual review of psychology, 67*, 439-463.

Rietveld, E. (2008a). Situated normativity: The normative aspect of embodied cognition in unreflective action. *Mind, 117*, 973-1001.

Rietveld, E. (2008b). Special section: The skillful body as a concernful system of possible actions phenomena and neurodynamics. *Theory & Psychology, 18*, 341-363.

Rietveld, E. (2008c). *Unreflective action. A philosophical contribution to integrative neuroscience*. Amsterdam: Institute for Logic, Language and Computation.

Rietveld, E. (2012). Bodily intentionality and social affordances in context. In F. Paglieri (Ed.), *Consciousness in interaction. The role of the natural and social context in shaping consciousness* (pp. 207-226). Amsterdam: J. Benjamins.

Rietveld, E., De Haan, S., & Denys, D. (2013). Social affordances in context: What is it that we are bodily responsive to? *Behavioral and Brain Sciences, 36*(4), 436.

Rietveld, E., & Kiverstein, J. (2014). A rich landscape of affordances. *Ecological Psychology, 26*(4), 325-352.

Robbins, J., & Rumsey, A. (2008). Introduction: Cultural and linguistic anthropology and the opacity of other minds. *Anthropological Quarterly, 81*(2), 407-420.

Roepstorff, A. (2013). Interactively human: Sharing time, constructing materiality. *Behavioral and Brain Sciences, 36*(3), 224-225.

Roepstorff, A., & Frith, C. (2004). What's at the top in the top-down control of action? Script-sharing and 'top-top'control of action in cognitive experiments. *Psychological Research, 68*(2-3), 189-198.

Roepstorff, A., Niewöhner, J., & Beck, S. (2010). Enculturing brains through patterned practices. *Neural Networks, 23*(8), 1051-1059.

Rogoff, B. (2003). *The cultural nature of human development*: Oxford university press.

Rumsey, A. (2013). Intersubjectivity, deception and the 'opacity of other minds': Perspectives from Highland New Guinea and beyond. *Language & Communication, 33*(3), 326-343.

Sacheli, L. M., Christensen, A., Giese, M. A., Taubert, N., Pavone, E. F., Aglioti, S. M., & Candidi, M. (2015). Prejudiced interactions: implicit racial bias reduces predictive simulation during joint action with an out-group avatar. *Scientific reports, 5*, 8507.

Salge, C., Glackin, C., & Polani, D. (2014). Changing the environment based on empowerment as intrinsic motivation. *Entropy, 16*(5), 2789-2819.

Satne, G. (2015). The social roots of normativity. *Phenomenology and the Cognitive Sciences, 14*(4), 673-682.

Scarantino, A. (2015). Information as a probabilistic difference maker. *Australasian Journal of Philosophy, 93*(3), 419-443.

Scarantino, A., & Piccinini, G. (2010). Information without truth. *Metaphilosophy, 41*(3), 313-330.

Schieffelin, B. B., & Ochs, E. (1986). *Language socialization across cultures*: Cambridge University Press.

Searle, J. (1991). Response: The background of intentionality and action. In E. L. a. R. v. Gulick (Ed.), *John Searle and his critics* (pp. 289-300). Oxford: Basil Blackwell.

Searle, J. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT press.

Searle, J. (1995). *The construction of social reality*: Simon and Schuster.

Searle, J. (2010). *Making the social world: The structure of human civilization*: Oxford University Press.

Seligman, R., & Kirmayer, L. J. (2008). Dissociative experience and cultural neuroscience: Narrative, metaphor and mechanism. *Culture, medicine and psychiatry, 32*(1), 31-64.

Sengupta, B., Tozzi, A., Cooray, G. K., Douglas, P. K., & Friston, K. J. (2016). Towards a neuronal gauge theory. *PLOS Biology, 14*(3).

Shagrir, O. (2006). Why we view the brain as a computer. *Synthese, 153*(3), 393-416.

Shagrir, O. (2010). Brains as analog-model computers. *Studies in History and Philosophy of Science Part A, 41*(3), 271-279.

Shapiro, L. (2010). *Embodied cognition*. New York: Routledge.

Silva, P., Garganta, J., Araújo, D., Davids, K., & Aguiar, P. (2013). Shared knowledge or shared affordances? Insights from an ecological dynamics approach to team coordination in sports. *Sports Medicine, 43*(9), 765-772.

Skinner, B. F. (2011). *About behaviorism*. New York: Vintage.

Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Oxford: Blackwell Publishers.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142): Harvard University Press Cambridge, MA.

Sporns, O. (2010). *Networks of the brain*. Cambridge: MA: MIT Press.

Stephan, K. E., Kasper, L., Harrison, L. M., Daunizeau, J., den Ouden, H. E., Breakspear, M., & Friston, K. J. (2008). Nonlinear dynamic causal models for fMRI. *Neuroimage, 42*(2), 649-662.

Sterelny, K. (2003). *Thought in a hostile world: The evolution of human cognition.* Oxford: Blackwell.

Sterelny, K. (2007). Social intelligence, human intelligence and niche construction. . *Philosophical transactions of the Royal Society B: Biological sciences, 362*, 719-730.

Sterelny, K. (2015). Content, control and display: The natural origins of content. . *Philosophia, 43*, 549-564.

Terrone, E., & Tagliafico, D. (2014). Normativity of the background: a contextualist account of social facts. In *Perspectives on Social Ontology and Social Cognition* (pp. 69-86): Springer.

Thompson, E. (2010). *Mind in life: Biology, phenomenology, and the sciences of mind*: Harvard University Press.

Thompson, E., & Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends in Cognitive Sciences, 5*(10), 418-425.

Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, MA: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*(5), 675-691.

Tschacher, W., & Haken, H. (2007). Intentionality in non-equilibrium systems? The functional aspects of self-organized pattern formation. *New Ideas in Psychology, 25*(1), 1-15.

Tuomela, R. (2007). *The philosophy of sociality: The shared point of view*: Oxford University Press.

Turvey, M. T. (1992). Affordances and prospective control: An outline of the ontology. *Ecological Psychology, 4*, 173-187.

Turvey, M. T., Shaw, R., Reed, E., & Mace, W. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn. *Cognition, 9*(237-304).

van Dijk, L., Withagen, R., & Bongers, R. M. (2015). Information without content: A Gibsonian reply to enactivists' worries. *Cognition, 134*, 210-214.

Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies, 3*(4), 330-349.

Varela, F. J. (1999). *Ethical know-how: Action, wisdom, and cognition*: Stanford University Press.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*: MIT press.

Veissière, S. (2016). Varieties of Tulpa experiences: the hypnotic nature of human sociality, personhood, and interphenomenality. *Hypnosis and meditation: Towards an integrative science of conscious planes*, 55-78.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological functions. In. Cambridge, MA: Harvard University Press.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological review, 20*(2), 158-177.

Whitehouse, H. (2002). Modes of religiosity: Towards a cognitive explanation of the sociopolitical dynamics of religion. *Method and theory in the study of religion, 14*(3-4), 293-315.

Whitehouse, H. (2004). *Modes of religiosity: A cognitive theory of religious transmission*. Oxford: Rowman Altamira.

Wilson, R. A., & Clark, A. (2009). How to situate cognition. Letting nature take its course. In P. Robbins & M. Ayede (Eds.), *The Cambridge handbook of situated cognition* (pp. 55-77). Cambridge: Cambridge University Press.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.

Zahavi, D., & Satne, G. (2015). Varieties of shared intentionality: Tomasello and classical phenomenology. In J. A. Bell, A. Cutrofello, & P. M. Livingston (Eds.), *Beyond the analytic-continental divide: Pluralist philosophy in the twenty-first century* (pp. 305-325). New York and London: Routledge.

Zawidzki, T. (2013). *Mindshaping: A new framework for understanding human social cognition*. Cambridge, MA: MIT Press.

8.2. Chapter 2: Answering Schrödinger's question: A free-energy formulation

Original publication details:

Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2018). Answering Schrödinger's question: A free-energy formulation. *Physics of Life Reviews*, 24: 1–16.
Target paper with 14 peer commentaries.
doi.org/10.1016/j.plrev.2017.09.001.

Authors:
Maxwell James Désormeau Ramstead[1,2*]
Paul Benjamin Badcock[3,4,5]
Karl John Friston[6†]

Affiliations:
[1]Department of Philosophy, McGill University, Montreal, Quebec, Canada.
[2]Division of Social and Transcultural Psychiatry, Department of Psychiatry, McGill University, Montreal, Quebec, Canada.
[3]Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, Australia, 3010.
[4]Centre for Youth Mental Health, The University of Melbourne, Melbourne, Australia, 3052.
[5]Orygen, the National Centre of Excellence in Youth Mental Health, Melbourne, Australia, 3052.
[6]Wellcome Trust Centre for Neuroimaging, University College London, London, UK, WC1N3BG.
*Correspondence to: maxwell.ramstead@mail.mcgill.ca.
†Senior author.

Abstract:
The free-energy principle (FEP) is a formal model of neuronal processes that is widely recognised in neuroscience as a unifying theory of the brain and biobehaviour. More recently, however, it has been extended beyond the brain to explain the dynamics of living systems, and

their unique capacity to avoid decay. The aim of this review is to synthesise these advances with a meta-theoretical ontology of biological systems called variational neuroethology, which integrates the FEP with Tinbergen's four research questions to explain biological systems across spatial and temporal scales. We exemplify this framework by applying it to Homo sapiens, before translating variational neuroethology into a systematic research heuristic that supplies the biological, cognitive, and social sciences with a computationally tractable guide to discovery.

Keywords:
Free energy principle; Complex adaptive systems; Evolutionary systems theory; Hierarchically mechanistic mind; Physics of the mind; Variational neuroethology

Highlights:

We describe a meta-theoretical ontology of life based on the free energy principle.
We propose a multiscale formulation of the free energy principle.
We translate our ontology into a systematic research heuristic for life sciences.
We apply this meta-theoretical ontology and research heuristic to Homo sapiens.

Acknowledgments:
Work on this article was supported by the Wellcome Trust (K. Friston) and the Social Sciences and Humanities Research Council of Canada (M. J. D. Ramstead). We thank Jelle Bruineberg, Nabil Bouizegarene, Axel Constant, Michael Kirchhoff, Laurence Kirmayer, Noah Moss-Brender, Jonathan St-Onge, Samuel Veissière, Ishan Walpola, and Julian Xue, for helpful discussions and comments on earlier versions of this paper. Finally, we would like to thank an anonymous reviewer for invaluable guidance in describing these ideas.

In the previous chapter,

I proposed a theoretical framework, called the cultural affordances framework, to model human social and cultural cognition as well as implicit cultural learning in human social groups. More specifically, the cultural affordances framework is a mechanistic theory of human enculturation and social cognition, the dynamics of which are formulated across spatial and temporal scales; i.e., the framework provides a model of the mechanisms, operative at various scales (from that of neuronal ensembles to that of cultural dynamics), in virtue of which human agents become encultured and acquire the cultural norms and ways of doing things that characterise their niche.

However, the cultural affordances model lacked a principled mathematical account of how the components of this multiscale mechanism interact, and remained mostly silent on the ways these components are integrated at and across the various spatial and temporal scales at which they exist. The purpose of this chapter is to provide precisely such an account of multiscale systematic integration.

1. Introduction

As Schrödinger (1944) famously observed many years ago, living systems are unique among natural systems because they appear to resist the second law of thermodynamics by persisting as bounded, self-organizing systems over time. How is this remarkable feat possible? What is life? How is it realised in physical systems? And how can we explain and predict its various manifestations and behaviours? By asking such questions, Schrödinger inspired a new line of inquiry that broadly centres on evolutionary systems theory (EST), which has become one of the most pervasive paradigms in modern science. Closely related to complexity science, EST is an interdisciplinary field that builds on the pioneering efforts of theorists like Fisher, Wright, Haldane and Prigogine (Ao, 2005; Fisher, 1930; Sarkar, 1992; Wright, 1932), and has found expression in influential models such as Eigen and Schuster's hypercycles (Eigen & Schuster, 1979). Put simply, EST explains dynamic, evolving systems in terms of the reciprocal relationship between general selection and self-organisation (Badcock, 2012; Depew & Weber, 1995; Kauffman, 1993).

Originating from biology, general selection entails three interacting principles of change: variation, selection, and retention (Caporael, 2001). This Darwinian process not only applies to organisms (i.e., natural selection), but acts on all dynamically coupled systems (e.g., molecules, neural synapses, behaviours, theories, and technologies; Blackmore, 2000; Cziko, 1995; Mesoudi, Whiten, & Laland, 2006), and is a universal principle that cuts across both statistical and quantum mechanics (Ao, 2008; Campbell, 2016). On the other hand, self-organisation stems from dynamic systems theory in physics (Haken, 1983; Nicolis & Prigogine, 1977; Prigogine & Stengers, 1984), and refers to the emergence of functional, higher-order patterns resulting from recursive interactions among the simpler components of coupled dynamical systems over time (see Box 1). To date, research in EST has focused on complex adaptive systems—such as the brain (Haken, 1996; Kelso, 1995), social systems (Miller & Page, 2009), and the biosphere (Levin, 1998)—which adapt to the environment through an autonomous process of selection that recruits the outcomes of locally interacting components—within that system—to select a subset for replication or enhancement (Levin, 2003).

At the turn of the millennium, the principles of EST inspired a new theory in neuroscience called the free-energy principle (FEP). Drawn chiefly from statistical thermodynamics and machine learning, the FEP is a formal model of neuronal processes that was initially proposed to explain perception, learning and action (Friston, 2005; Friston, Kilner, & Harrison, 2006b), but has since been extended to explain the evolution, development, form,

and function of the brain (Friston, 2010). More recently, it has also been applied to biological systems across spatial and temporal scales, ranging from phenomena at the micro-scale (e.g., dendritic self-organisation and morphogenesis; Friston, Levin, Sengupta, & Pezzulo, 2015; Kiebel & Friston, 2011), across intermediate scales (e.g., cultural ensembles; Ramstead, Veissière, & Kirmayer, 2016), and at the macro-scale (e.g., natural selection; Friston & Ao, 2011). We believe that this theory puts us in a strong position to answer Schrödinger's question, and at the same time, shed new light on the mind, body, behaviour, and society.

We begin by describing how the FEP offers a plausible, mechanistic EST that applies to living systems in general. We then combine this variational principle with Tinbergen's four research questions in biology to describe a new scientific ontology—called variational neuroethology—that can be used to develop mathematically tractable, substantive explanations of living organisms. We conclude by applying this meta-theory to the most complex living system known to date—namely, ourselves—before translating this framework into a systematic research heuristic. In doing so, we hope to highlight a plausible, computationally tractable guide to discovery in the biological, cognitive and social sciences.

2. The free energy formulation

The FEP is a mathematical formulation that explains, from first principles, the characteristics of biological systems that are able to resist decay and persist over time. It rests on the idea that all biological systems instantiate a hierarchical generative model of the world that implicitly minimises its internal entropy by minimising free energy. In virtue of the self-organisation inherent in nonequilibrium steady-state, systems will apparently violate the second law of thermodynamics. See Ao, Chen, and Shi (2013) for an interesting treatment of relative entropy in this context (that does not require detailed balance assumptions). From our perspective, this sort of behaviour can be cast in terms of a dynamics (i.e., conservative flow) that appears to minimise a variational free energy, which constitutes an upper bound on the entropy of a system's Markov blanket (see Friston (2013) and Box 2). Technically, free energy is an information theoretic quantity that limits (by being greater than) the entropy of sensory exchanges between a biotic system (e.g., the brain) and the environment. A generative model is a probabilistic mapping from causes in the environment to observed consequences (e.g., sensory data); while entropy refers to the (long-term) average of surprise (or surprisal)—the negative log probability of sensory samples encountered by an agent (Friston, 2010). Under this formalism, for an organism to resist dissipation and persist as an adaptive system that is part of, coupled with, and yet statistically independent from, the larger system in which it is

embedded, it must embody a probabilistic model of the statistical interdependencies and regularities of its environment. We elaborate on this next.

2.1. Living systems, ergodicity, and phenotypes

All biological systems exhibit a specific form of self-organisation, which has been sculpted by natural selection to allow them to actively maintain their integrity by revisiting characteristic states within well-defined bounds of their conceivable phase spaces (see Box 1). In other words, there is a high probability that an organism will occupy a relatively small, bounded set of states—its viability set (Di Paolo & Thompson, 2014)—within the total set of possible states that it might occupy (i.e., its phase space). In terms of information theory, this means that the probability density function that describes the possible states of the system has low entropy. So, how do living systems perform this feat?

This is simpler than it might seem, and rests on the fact that all living systems revisit a bounded set of states repeatedly (i.e., they are locally ergodic). At every scale—from the oscillations of neuronal activity over milliseconds, through to the pulsations of our heart and our daily routines—we find ourselves in similar states of mind and body. This is the remarkable fact about living systems. All other self-organising systems, from snowflakes to solar systems, follow an inevitable and irreversible path to disorder. Conversely, biological systems are characterised by a random dynamical attractor—a set of attracting states that are frequently revisited. Indeed, the characteristics by which we define living systems are simply statements about the characteristic, attracting states in which we find them (Friston, 2013). This set of attracting states can be interpreted as the extended phenotype of the organism—its morphology, physiology, behavioural patterns, cultural patterns, and designer environments (Clark, 2013). This conception of the extended phenotype as the set of attracting states of a coupled dynamical system is supported by evidence from simulation studies of morphogenesis, (e.g., Friston et al., 2015). Further supportive evidence comes from studies of cancer genesis and progression, where the success of approaches employing endogenous networks provides a striking example of employing statistical methods (the Markov blanket formalism) to separate internal (phenotypical) states from external ones (Yuan, Zhu, Wang, Li, & Ao, 2017). This conception of the topology of the phase space is supported by recent work on early myelopoiesis in real biological systems as well (Su et al., 2017). In this study, the core molecular endogenous network under consideration was cast as a set of dynamical equations, yielding structurally robust states that can be interpreted in relation to known cellular phenotypes.

The implications of this are profound. It means that all biotic agents move, systematically, towards attracting states (i.e., those with high probability) to counter the dispersive effects of random fluctuations. Consequently, any living system will appear, on average, to move up the probability gradients that define its attracting set—and the very characteristics responsible for its existence. Thus, living systems do not just destroy energy gradients (by gravitating towards free energy minima), they also create and maintain them by climbing the probability gradients that surround such extrema. In other words, living systems carve out and inhabit minima in free energy landscapes, precluding the dissipation of their states over phase space. This (nonequilibrium steady-state) behaviour differentiates living states from other states, like decay and death (Bruineberg & Rietveld, 2014; Jarzynski, 1997; Schrödinger, 1944; Tomé, 2006). Technically, this gradient-building behaviour can be expressed as the flow over a landscape that corresponds to the log probability of any state being occupied. This probability is also known as 'Bayesian model evidence' (Friston, 2010). This means living systems are effectively self-evidencing—they move to maximise the evidence of their existence (Hohwy, 2016). So how do they achieve this?

This is where the FEP comes in. It asserts that all biological systems maintain their integrity by actively reducing the disorder or dispersion (i.e., entropy) of their sensory and physiological states by minimising their variational free energy (Friston (2010); see Figure 1 and Box 2). Because the repertoire of functional or adaptive states occupied by an organism is limited, the probability distribution over these characteristic states has low entropy: there is a high probability the organism will revisit a small number of states. Thus, an organism's distal imperative of survival and maintaining functional states within physiological bounds (i.e., homeostasis and allostasis) translates into a proximal avoidance of surprise (Friston, 2010). Although surprise itself cannot be evaluated, since free energy imposes an upper bound on surprise, biological systems can minimise surprise by minimising their variational free energy. From the point of view of a physicist, surprise corresponds to thermodynamic potential energy (Seifert, 2012), such that minimising (the average) variational free energy entails the minimisation of thermodynamic entropy. Consistent with EST, this propensity to minimise surprise is the result of natural selection (that itself can be seen as a free energy minimising process; see below)—self-organising systems that are able to avoid entropic, internal phase-transitions have been selected over those that could not (Friston et al., 2006b). This begs the question of how biological systems minimise free energy. To answer this, we will now take a closer look at the relation between surprise and (variational) free energy by introducing the notion of Markov blankets, which is central to the variational neuroethology described later.

2.2. Free energy, surprise, and Markov blankets

To understand the subtle but important difference between surprise and free energy, we have to look more carefully at what constitutes a system. Clearly, one needs to differentiate between the system and its environment—those states that constitute or are intrinsic to the system and those that are not. To do this, we have to introduce a third set of states that separates internal from external states. This is known as a Markov blanket. Markov blankets establish a conditional independence between internal and external states that renders the inside open to the outside, but only in a conditional sense (i.e., the internal states only 'see' the external states through the 'veil' of the Markov blanket; Clark, 2017; Friston, 2013). The Markov blanket can be further divided into 'sensory' and 'active' states that are distinguished in the following way: internal states cannot influence sensory states, while external states cannot influence active states (Friston, 2012). With these conditional independencies in place, we now have a well-defined (statistical) separation between the internal and external states of any system. A Markov blanket can be thought of as the surface of a cell, the states of our sensory epithelia, or carefully chosen nodes of the World Wide Web surrounding a particular province.

With Markov blankets in mind, it is fairly straightforward to show that the internal states must—by virtue of minimising surprise—encode a probability distribution over the external states; namely, the causes of sensory impressions on the Markov blanket. This brings us back to free energy. Free energy is a functional (i.e., the function of a function) that describes the probability distribution encoded by the internal states of the Markov blanket. Note that this is different from surprise, which is a function of the states of the Markov blanket itself. In other words, free energy is a function of probabilistic beliefs, encoded by internal states about external states (i.e., expectations about the probable causes of sensory input). When these beliefs are equal to the posterior probability over external states, free energy becomes equivalent to surprise. Otherwise, it is always slightly greater than (i.e., imposes an upper bound on) surprise. This means that living systems can be characterised as minimising variational free energy, and therefore surprise, where the minimisation of variational free energy entails the optimisation of beliefs about things beyond or behind the Markov blanket (see Box 2). This inferential aspect leads us to notions like embodied inference (Bruineberg, Kiverstein, & Rietveld, 2016; Gallagher & Allen, 2016), the Bayesian brain (Knill & Pouget, 2004), the Bayesian cell (Kiebel & Friston, 2011), and even a Bayesian culture (Friston & Frith, 2015; Ramstead et al., 2016).

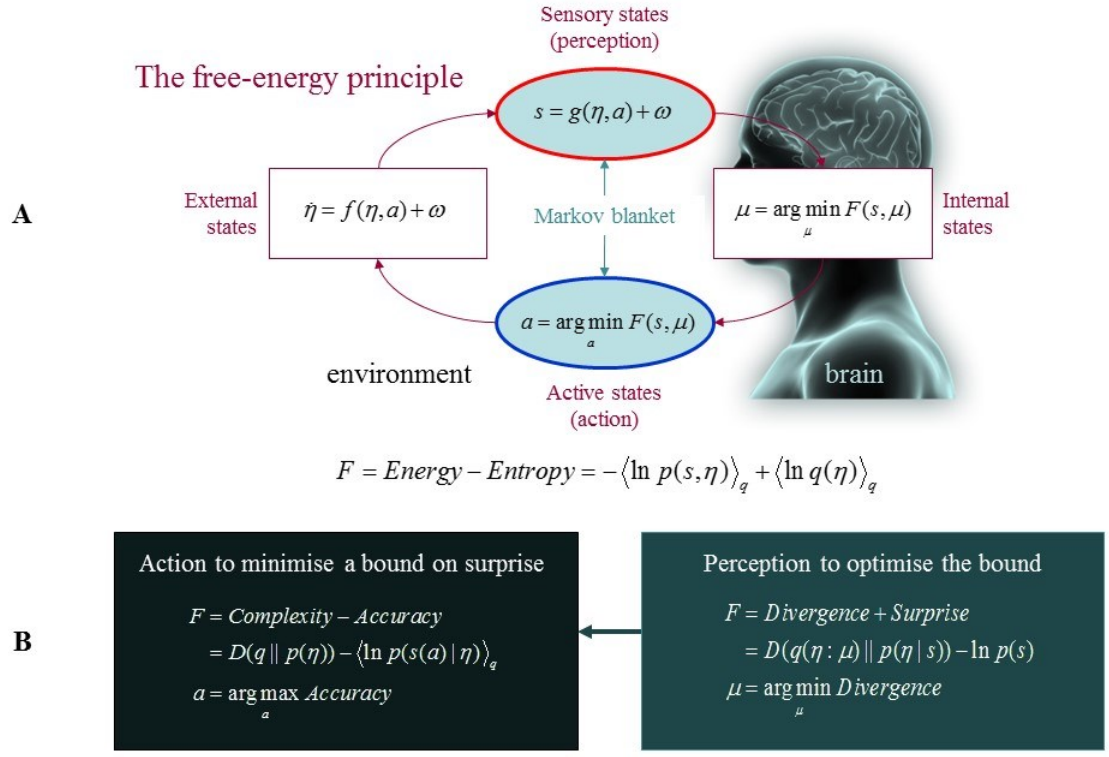


Fig. 1. *The free energy principle.* (**A**) Schematic of the quantities that define free-energy. These include the internal states of a system μ (e.g. a brain) and quantities describing exchange with the world; namely, sensory input $s = g(\eta, a) + \omega$ and action a that changes the way the environment is sampled. The environment is described by equations of motion, $\dot{\eta}$ $) + \omega$, that specify the dynamics of (hidden) states of the world η. Here, ω denote random fluctuations. Internal states and action both change to minimise free-energy, which is a function of sensory input and a probabilistic representation (variational density) $q(\eta : \mu)$ encoded by the internal states. (**B**): Alternative expressions for the free-energy illustrating what its minimisation entails. For action, free-energy can only be suppressed by increasing the accuracy of sensory data (i.e., selectively sampling data that are predicted). Conversely, optimising internal states make the representation an approximate conditional density on the causes of sensory input (by minimising divergence). This optimisation makes the free-energy bound on surprise tighter and enables action to avoid surprising sensations.

In short, then, how do Markov blankets relate to the FEP? The FEP tells us how the quantities that define Markov blankets change as the system moves towards its variational free energy minimum (following Hamilton's principle of least action; see Figure 1 and Box 2). It asserts that for a system to resist entropic erosion and maintain itself in a bounded set of states (i.e., to possess a generalised homeostasis), it must instantiate a causal, statistical model of its eco-niche relation (Conant & Ross Ashby, 1970; Friston, 2013; Friston, 2010; Seth, 2014). In

other words, an organism does not just encode a model of the world, it *is* a model of the world—a physical transcription of causal regularities in its eco-niche that has been sculpted by reciprocal interactions between self-organisation and selection over time. On the basis of these distinctions, we turn next to defining a fully generalizable ontology for biological systems based on a multiscale free energy formulation, which we call 'variational neuroethology'.

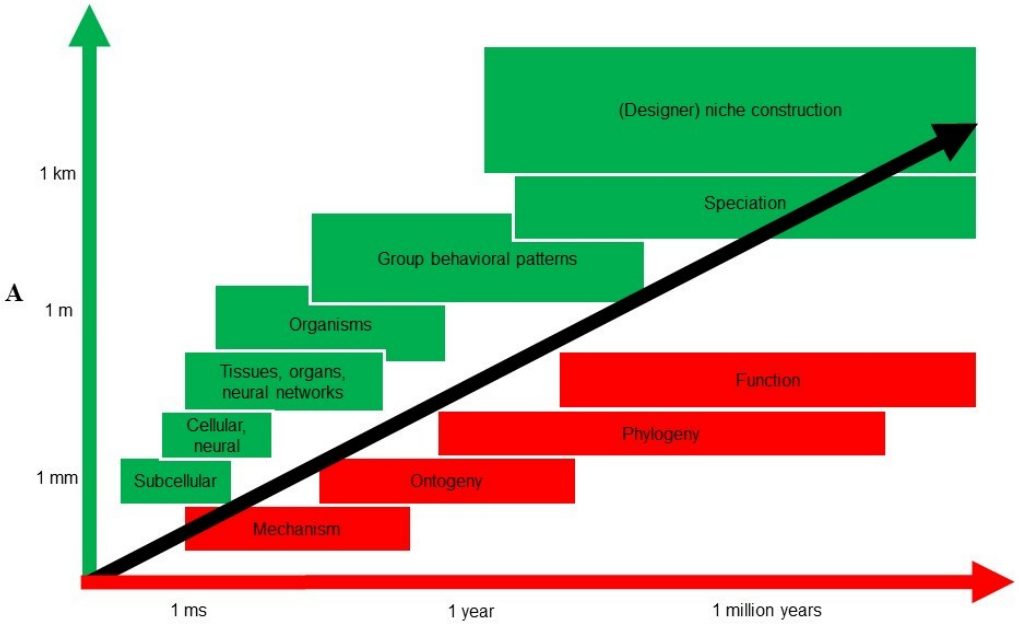
3. The big picture: A multiscale free energy formulation

The crux of our argument is that organisms can be described in terms of a (high dimensional) phase space induced by hierarchically nested Markov blankets. In other words, our ontology comprises populations of both spatially and temporally nested Markov blankets that occupy hierarchically nested regions in the total phase space of living systems. This sort of hierarchical organisation is a direct corollary of EST: since specific, functional (global) patterns of interacting (local) components need to be selected over competing alternatives to allow different levels of (informational, physical, chemical, biological, psychological, and sociocultural) organisation to emerge, the hierarchical nesting of Markov blankets instantiates Darwinian dynamics, which follows the same laws of statistical (or, strictly speaking, stochastic) thermodynamics, but in a nonequilibrium context that leads to self-organisation, self-assembly, and selective dynamics (Ao et al., 2013; Badcock, 2012; Friston, 2013; Martyushev & Seleznev, 2006).

3.1. Nested Markov blankets

To picture such hierarchical dynamics, it is useful to introduce the notion of a scale space. Scale spaces allow us to observe structures at different spatial scales. Imagine that you took a photograph, and then focused in progressively to examine smaller details. As you zoom in, you traverse a (spatial) scale space. The notion of a scale space is useful because the increase in scale, as we move from one hierarchical level of Markov blankets to the next, necessarily entails an increase in spatial scale. However, what were purposeful (i.e., free energy minimising) fluctuations at one scale now become fast random fluctuations at the next. This means that there is a concomitant increase in temporal scale as we ascend the spatial hierarchy. This composition of temporal and spatial scales is evident in the hierarchical organisation of the brain (Jung, Hwang, & Tani, 2015; Kiebel, Daunizeau, & Friston, 2008), and more broadly, suggests that self-organisation should occupy a limited domain (along the diagonal) of a scale space with spatial and temporal dimensions (Haken, 1983) (see Figure 2). Note that the use of a scale space is purely for descriptive purposes. The underlying system in question does not

change—just its level of description, the way it is measured, or the perspective taken on its hierarchical self-organisation.



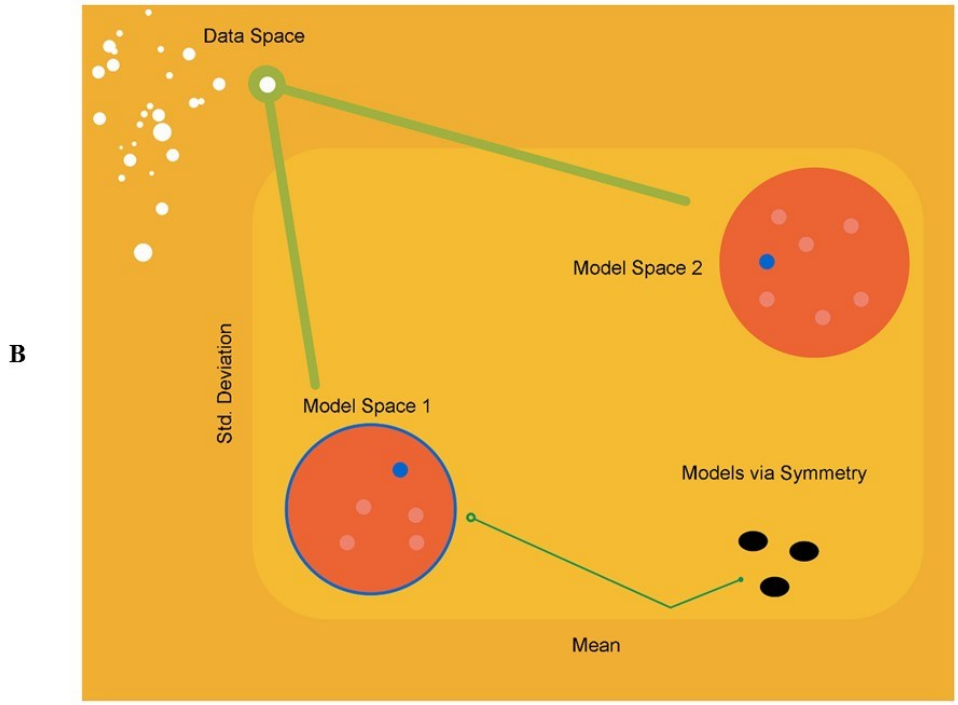


Fig. 2. Variational neuroethology. (**A**) The meta-theoretical ontology we propose, called 'variational neuroethology', uses the FEP to explain and predict how living systems instantiate adaptive free energy minimisation. We have indicated some scales at which free energy minimizing dynamics unfold. Since spatial and temporal scales are intrinsically correlated (i.e., events unfolding over long distances usually take more time to unfold), what we have is a scale space that is populated mostly along its diagonal (Bak, Tang, & Wiesenfeld, 1988; Dorogovtsev & Mendes, 2002; Mantegna & Stanley, 1995). (**B**) Equivalence classes of variational free energy minimizing systems. The free energy minimising dynamics at play are implemented by different kinds of mechanisms in different individual organisms and species, as a function of the coupling between their evolved phenotypes and biobehavioural patterns and the niches they inhabit and the scales under scrutiny. The gauge theoretical formalism for the FEP (Sengupta, Tozzi, Cooray, Douglas, & Friston, 2016) allows us to computationally model the regions of the biotic phase space, along its diagonal, that are apt to realize equivalent classes of dynamics. From Sengupta et al. (2016).

Recall from above that every ergodic system must possess an (ergodic) Markov blanket (Friston, 2013). This simple observation delivers us to the core of our argument. Thus, we should be able to describe the universe in terms of Markov blankets of Markov blankets—and Markov blankets all the way up, and all the way down (see Figure 3 and Box 3). To unpack this, consider an ensemble of cells, each equipped with a Markov blanket that corresponds to the cell surface. Because the internal states of each cell are sequestered behind their respective Markov blankets, all interactions between cells must be mediated by their Markov blankets. This means that we can describe the self-organisation of the cellular ensemble purely in terms of transactions among the (sensory and active) states of Markov blankets. However, these exchanges will themselves have a sparsity structure that induces a Markov blanket of Markov blankets. For example, one subset of the ensemble could be an organ that is surrounded by epithelia that provide a Markov blanket for all of the cells that constitute the internal states of the organ. However, this still follows exactly the same (statistical) structure—an ensemble of Markov blankets. We can then repeat this process, at increasingly larger scales of self-organisation, to create a series of hierarchically nested systems (e.g., the body) (Clark, 2017).

This sort of hierarchical structure provides a universal and recursive perspective to understand self-organisation across spatial and temporal scales, and to explain how each level contextualises (constrains) the levels both above and below. The hierarchal composition of Markov blankets within Markov blankets follows naturally from the existence of a Markov blanket that, in turn, is mandated by the existence of any system that can be distinguished from its external milieu. The key point here is that at every level, the same variational, surprise-reducing dynamics must be in play to supply Markov blankets for the level above. As we argue below, this idea offers a promising new research heuristic for the biological sciences.

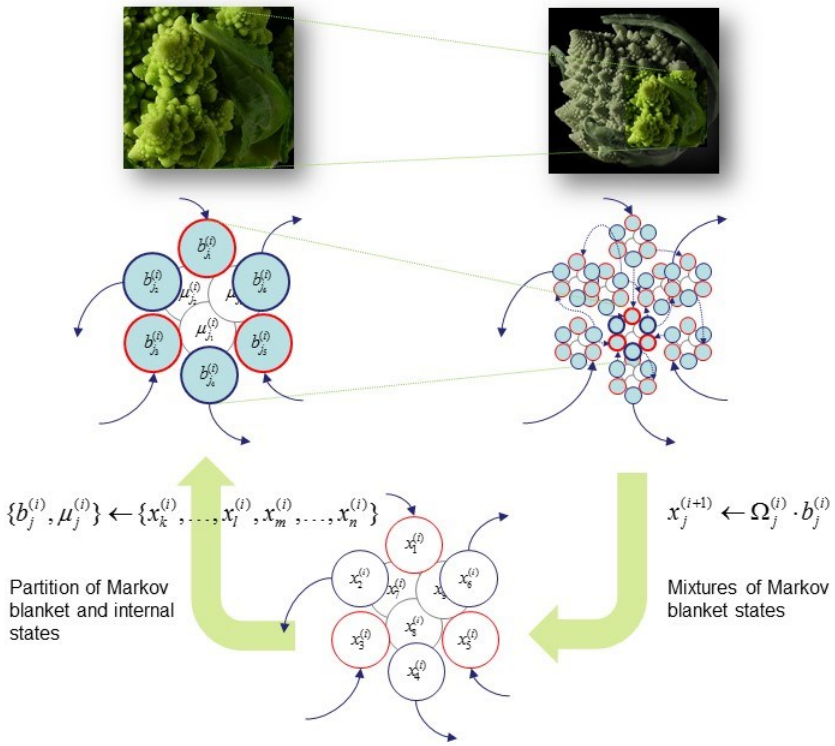


Fig. 3. Nested Markov blankets. This schematic illustrates the hierarchical construction of (scale-free) compositions of Markov blankets of Markov blankets. The idea here is that particles, cells or subsystems at one scale (each comprising a Markov blanket $b_j^{(i)}$ that enshrouds internal states $\mu_j^{(i)}$) constitute an ensemble of states with a sparse dependency structure, which induces a Markov blanket at the supraordinate scale. This allows one to construct Markov blankets of Markov blankets by (i) partitioning the states at one level into a series of internal subsets and their Markov blankets and (ii) creating states for the next level by taking mixtures of Markov blanket states. Note that the internal states can be ignored when going from one level to the next because they are conditionally independent of external states (i.e., mixtures of Markov blankets from other subsystems). The mixtures can be regarded as slow (unstable) modes that are referred to as *order parameters* in synergetics (Haken, 1983). Filled (cyan) circles correspond to Markov blanket states at the i-th scale, where, as in Figure 1, red denote sensory states and blue active states. The pictures (of Broccoli) in the upper panels illustrate the self similarity of this recursive partitioning.

3.2. Multiscale integration and variational neuroethology

Above, we suggested that living systems can be described in terms of hierarchically nested Markov blankets. The hierarchical interdependencies between Markov blankets provide a context for biological phenomena at every scale, which implies that the preservation of Markov blankets at a lower scale are necessary for the ongoing conditional independencies that form

the basis of Markov blankets at a higher scale, and vice-versa (see Figure 3 and Box 3). Put simply, the existence of every level depends upon every other level. Of particular interest here, however, is the importance of the different temporal scales that transcend spatial scales. Strictly speaking, free energy is only ever minimised diachronically—that is, over some discrete time span—as a process (Kirchhoff, 2015). After all, the FEP is a formulation of the constraints imposed on system dynamics: it tells us about the dynamics of living systems that obtain precisely because of what it means for physical systems to be alive (Friston, 2012).

The hierarchical free energy formulation we propose here encompasses the ways in which the FEP applies to each hierarchical level of organisation by articulating the minimisation of free energy across scales. Any system's operation or functioning (i.e., a particular, time extensive event in our ontology) can be seen as unfolding across spatial scales, driven by temporally integrated free energy minimisation dynamics. Because the time average of free energy is a Hamiltonian action, all we are saying here is that Hamilton's principle of least action (perhaps the most celebrated variational principle) is realised by hierarchically nested biophysical processes. Having said this, the assertion that the time average of free energy is a Hamiltonian action is non-trivial for general stochastic dynamics, and is still a subject of research, i.e., Tang, Yuan, and Ao (2014). Over any given time lapse, free energy minimisation can thus be framed as a hierarchical 'unpacking', across spatial scales, of the same invariant (or scale-free) dynamics. Since spatial and temporal scales are intrinsically correlated (i.e., events unfolding over long distances usually take more time to unfold (Bak et al., 1988; Dorogovtsev & Mendes, 2002; Mantegna & Stanley, 1995), what we have is a scale space that is populated mostly along its diagonal (see Figure 2).

Framing things in this way allows us to derive a computationally tractable, dynamical typology of the scientifically relevant events (types of biotic systems) at each timescale, and also to distinguish between different regions in the biophysical phase space, which roughly correspond with the ontologies of scientific disciplines concerned with different living systems (e.g., biology, psychology and the social sciences; Henriques, 2011). These self-organising, variational imperatives remain identical at each level of organisation, but differ fundamentally in how they are mediated. For example, the brain uses some form of belief propagation or predictive coding to maintain its Markov blanket (Dayan, Hinton, Neal, & Zemel, 1995; Rao & Ballard, 1999; Srinivasan, Laughlin, & Dubs, 1982), while evolution uses Bayesian model selection and stochastic sampling schemes (Ao, 2009; Campbell, 2016; Frank, 2012; Harper, 2011) to preserve the Markov blankets that underlie speciation at evolutionary timescales.

The notion of Markov blankets allows us to define a formal ontology of living systems. Markov blankets are nested within Markov blankets, over spatial and temporal scales. This suggests that there has to be an internal (scale-free or scale-invariant) consistency, in the sense that the variational free energy associated with a global Markov blanket (e.g., a species) has to conform to the same principles as all of its constituent Markov blankets (e.g., phenotypes), which applies recursively right down to the level of biological macromolecules. In this vein, the FEP has recently been formulated using the resources of gauge theory (Sengupta et al., 2016); see also the treatment of dynamic systems in Ao (2003). Gauge theory is a family of mathematical models that have been broadly applied in physics. The gauge theoretical formalism is used to define a Lagrangian, a functional that summarises the dynamics of a given system and preserves its global symmetry (see Box 4). This symmetry is broken by local forces, invoking a gauge field that restores the system's symmetry by compensating for these local perturbations. The FEP has been proposed as the Lagrangian of a gauge theory for living systems (Sengupta et al., 2016). Over any given time scale, free energy is minimised across every spatial scale, while local dynamics at each of these scales—determined by local Markov blanket features—perturb the Lagrangian. However, these perturbations are then compensated for by one or more gauge fields, which are introduced when an organism either changes its internal (e.g., perceptual) states or acts upon the world to minimise surprise. The upshot of this is that it allows us to computationally model scale-invariant free energy minimisation dynamics across temporal and spatial scales. As we ascend nested Markov blankets, and as we consider different biological systems, free energy is minimised in dynamically equivalent, but mechanistically heterogeneous, ways.

This brings us to a meta-theoretical ontology of biological systems derived from the FEP—variational neuroethology—that can be used to explain and predict how living systems, at any spatial and temporal scale, instantiate the dynamics of adaptive free energy minimisation. The precise nature of free energy minimising mechanisms will vary from organism to organism and species to species, as a function of their evolved phenotypes and biobehavioural patterns. In other words, the FEP supplies a universal Lagrangian that extends across all spatial and temporal scales, but the particular ways in which it is implemented will vary according to the species, organisms, and scales under scrutiny. This allows us to computationally model the regions of the phase space populated by different organisms—along its diagonal—that realise equivalent classes of dynamics, and to recast the issue of ecological problem-solving as the problem of finding local free energy minima. At first glance, analysing such complex dynamics across time and space might seem intractable, but conveniently,

biologists have long been familiar with a highly compatible framework that allows us to translate our variational neuroethology into viable scientific practice.

4. Integrating the multiscale free energy formulation with Tinbergen's four questions

In the sciences of life, mind, and society, using free energy minimisation—realised by nested Markov blankets—enables us to explain dynamics both at and between nested scale spaces. However, although the FEP supplies a powerful, mechanistic theory that captures biological dynamics across spatial and temporal scales, it can only offer partial insight into the particular features of a given species, which instantiate distinct, embodied models of specific adaptive needs and environmental niches (Allen & Friston, 2016; Clark, 2013, 2015). In his treatment of Darwinian dynamics, Ao (2005) describes how a complete explanation for any biological system requires two types of laws: those that account for the structure of evolutionary dynamics (e.g., natural selection); and those that can explain the structure of each dynamical component (e.g., genes). We suggest here that the FEP is analogous to the first type of law—it describes, formally, the (entropy bounding) dynamics of all living systems. But what of the second type? Although it is highly generalizable, the explanatory scope of the FEP is limited—it only imposes relatively modest constraints on the classes of dynamical patterns (i.e., complex adaptive systems) that count as living. The FEP therefore requires a complementary evolutionary (i.e., ultimate) account that explains the specific adaptive solutions responsible for producing different embodied models, along with the proximate processes that produce every phenotype. Again, this appeals to the importance of timescales when describing the dynamic ways in which living systems minimise free energy across space and time.

Fortunately, the importance of temporal scales has long been recognised in biology, particularly after Tinbergen (1963) proposed his four key levels of biological explanation (i.e., adaptation, phylogeny, ontogeny, and mechanism). Given the success of this explanatory framework in biology, we suggest that Tinbergen's levels of inquiry might be apt to elucidate structural laws that supplement the general principles provided by the FEP. Clearly, Tinbergen's 'four questions' have a long history of producing detailed explanations for the evolution, development, form and function of different species, and by focusing on different temporal scales (ranging from a species' evolution to an organism's behaviour in real-time), they allow us to capture the complexities of every corresponding spatial scale (from sub-atomic particles, atoms, and cells, all the way up). Conversely, the FEP furnishes a biologically plausible EST that applies both to ultimate and proximate processes, supplying a universal principle that intersects all four of Tinbergen's levels of inquiry. We believe, then, that these

two paradigms are highly commensurate—the FEP describes a biological imperative to model the world that constrains the dynamics of all living systems (at any time-scale), while Tinbergen has offered a distinctive but complementary framework that allows us to develop substantive explanations for the phenotypic traits and behaviours of any given species or organism. Accordingly, the variational neuroethology proposed here hinges on their synthesis. To exemplify this approach, we turn now to its recent application to the world's most complex living system—us.

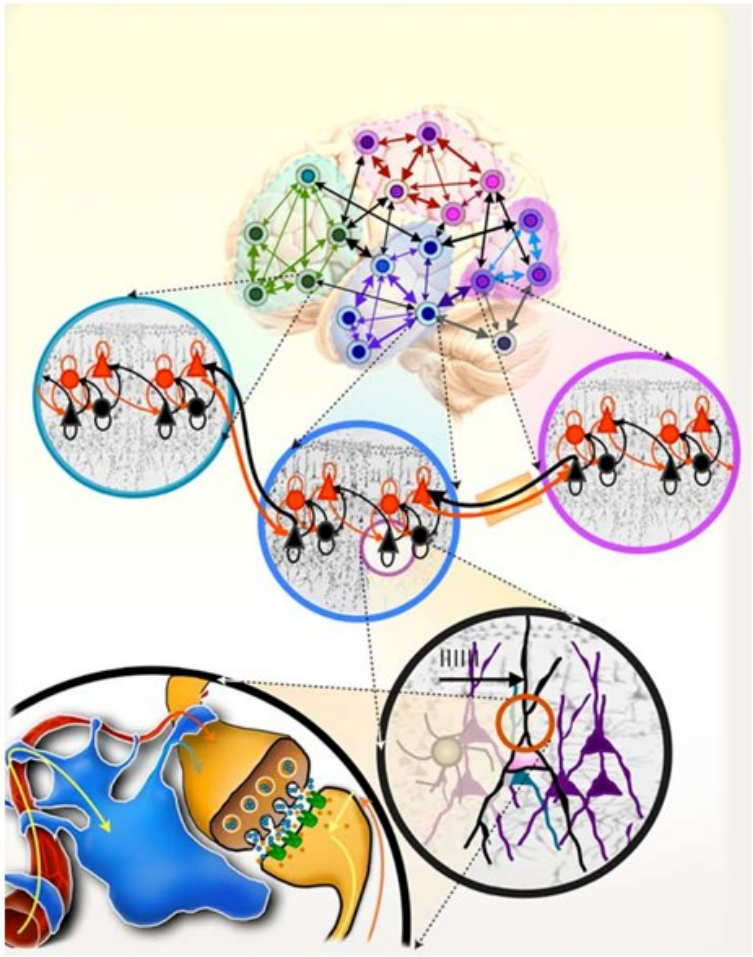
4.1. Variational neuroethology applied: Hierarchically mechanistic minds

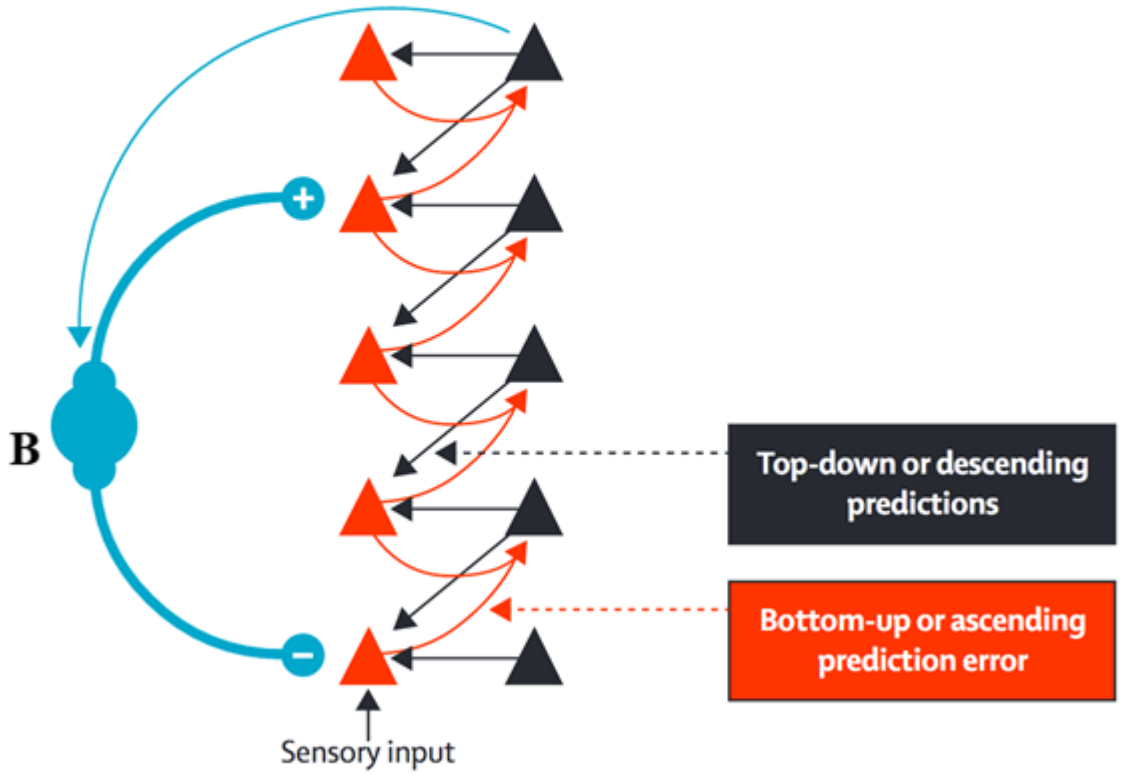
The FEP is best known in neuroscience, where it has been used to explain the structure, function, and dynamics of the brain. In this context, the FEP concords with predictive coding by describing the brain as a hierarchical 'inference machine' that minimises prediction error by seeking to match incoming sensory inputs with top-down (neuronally-encoded) predictions (Clark, 2013, 2015; Friston, 2010; Hohwy, 2016). This occurs in two ways: we can either improve our predictions by altering internal states (i.e., perception); or we can act upon the world to confirm our predictions (i.e., action). Thus, action and perception operate synergistically to optimise an organism's (Bayesian) model of the environment. As discussed, the FEP also transcends predictive coding by extending beyond the brain to explain behaviour, the phenotype, and all other biotic phenomena that span evolutionary timescales and spatially distributed ensembles. However, to understand the particular features of the human brain, the FEP requires recourse to research in psychology and the social sciences (e.g., evolutionary and cognitive anthropology), which target the specific ecobiopsychosocial processes responsible for the embodied models and behaviour of Homo sapiens in particular.

To address this need, an interdisciplinary EST of the embodied brain has recently been forwarded called the hierarchically mechanistic mind (HMM). Initially proposed to unify evolutionary and developmental psychology, the HMM is an evidence-based model of neurocognition and biobehaviour that synthesises the FEP with major paradigms in psychology to situate the brain within the broader evolutionary, developmental, and real-time processes that produce human behaviour, phenotypes, and niches (Badcock, 2012; Badcock, Davey, Whittle, Allen, & Friston, 2017). Specifically, this model defines the human brain as an embodied, complex adaptive system that adaptively minimises the entropy of its internal (i.e., sensory and physiological) states through recursive interactions between hierarchically organized, functionally differentiated neural subsystems (Badcock, 2012). This hierarchy ranges from lower-order, highly segregated neurocognitive systems responsible for

sensorimotor processing, through to the highly integrated association areas that underlie the sophisticated cognitive faculties unique to humans (Badcock et al., 2017). The HMM resonates with structural and functional imaging studies in network neuroscience, which show that the brain entails a multiscale hierarchical organisation characterised by the repeated encapsulation of smaller neural elements in larger ones (Kaiser, Hilgetag, & Kötter, 2010; Park & Friston, 2013; also see Figure 4, panel A). Predictive coding approaches suggest that this sort of architecture entails a hierarchical generative model that minimises prediction error via recurrent message-passing between cortical levels (Figure 4, panel B), affording a neurobiologically plausible, mechanistic theory of the functional integration of anatomically segregated neural networks (Bastos et al., 2012; Park & Friston, 2013; Shipp, 2016).

A




B
Top-down or descending predictions
Bottom-up or ascending prediction error
Sensory input
Prediction error (superficial pyramidal cells)
Posterior expectations (deep pyramidal cells)

C

Process **Temporal scale**	**Level of inquiry** **State space**	**Free energy formulation** **Psychological paradigm(s)**
Mechanism (real-time)	*Neurocognition* Perception & action + Learning & attention The individual in context	$\mu_x^{(i)} = \arg\min F\left(\tilde{s}(a), \mu^{(i)} \middle| m^{(i)}\right)$ $\mu_a^{(i)} = \arg\min F\left(\tilde{s}(a), \mu^{(i)} \middle| m^{(i)}\right)$ $\mu_\gamma^{(i)} = \arg\min \int dt\ F\left(\tilde{s}^{(i)}, \mu^{(i)} \middle| m^{(i)}\right)$ $\mu_\theta^{(i)} = \arg\min \int dt\ F\left(\tilde{s}^{(i)}, \mu^{(i)} \middle| m^{(i)}\right)$ Psychological subdisciplines
Ontogeny (developmental time)	*Neurodevelopment* The individual	$m^{(i)} = \arg\min \int dt\ F\left(\tilde{s}^{(i)}, \mu^{(i)} \middle| m^{(i)}\right)$ Developmental psychology
Phylogeny (intergenerational time)	*Neurophylogeny* Groups (e.g., kin)	$s = \arg\min \sum_{m^{(i)} \in s} \int dt\ \ F\left(\tilde{s}^{(i)}, \mu^{(i)} \middle| m^{(i)}\right)$ Evolutionary developmental biology and psychology
Adaptation (evolutionary time)	*Neural evolution* Homo sapiens	$c = \arg\min \sum_{m^{(i)} \in c} \int dt\ \ F\left(\tilde{s}^{(i)}, \mu^{(i)} \middle| m^{(i)}\right)$ Evolutionary psychology

Informational exchange

Fig. 4. The hierarchically mechanistic mind. (**A**) Schematic of the multiscale hierarchical organisation of brain networks. Neural networks are composed of nodes and their connections, called edges. A node, defined as an interacting unit of a network, is itself a network composed of smaller nodes interacting at a lower hierarchical level, producing nested neural networks that extend from neurons, to macrocolumns, and to macroscopic brain regions. From Park and Friston (2013). (**B**) A simple cortical hierarchy with ascending prediction errors and descending predictions. Superficial pyramidal cells (red triangles) compare expectations (at each level) with top-

down predictions from deep pyramidal cells (black triangles) at higher levels. Neuromodulatory gating or gain control (blue) of superficial pyramidal cells determines their relative influence on the deep pyramidal cells that encode expectations by modulating their precision. From Friston, Stephan, Montague, and Dolan (2014). (**C**): The variational neuroethology of human cognition and biobehaviour. $F(\tilde{\ }\ \ m^{(i)})$ represents the free-energy of the sensory data (over time), $\tilde{\ }$, and the states μ of an agent $m^{(i)} \in s$ that belongs to a subgroup $s \in c$ of class c. Action (a) governs the sampling of sensory data, and the physical states of the phenotype (μ) encode beliefs or expectations (and expectations of the mean of a probability distribution). Free energy minimisation dynamics vary across timescales, ranging from neurocognition in real-time (i.e., perception and action; learning and attention); neurodevelopment throughout the lifespan; epigenetic mechanisms that minimise free energy across generations (e.g., kin); and the process of adaptation, which involves the optimisation of human generative models over time and conspecifics via the inheritance of adaptive priors (Friston, 2011). These temporal processes are captured by Tinbergen's four levels of inquiry, which appeal to a dynamic causal hierarchy that is encapsulated by complementary paradigms in psychology: evolutionary psychology; evolutionary developmental approaches; developmental psychology; and the psychological subdisciplines (Badcock, 2012; Badcock et al., 2017)

Following EST, the hierarchical brain also exemplifies the complementary relationship between evolution and development—selection has canalized early sensorimotor regions that serve as neurodevelopmental anchors, allowing for the protracted, activity-dependent self-organisation of 'domain-general' association cortices throughout development that enhance evolvability by responding flexibly to environmental change (Anderson & Finlay, 2014; Buckner & Krienen, 2013). Consistent with this, simulation studies of evolving networks have shown that a hierarchical neural structure enhances evolvability by adapting faster to new environments than non-hierarchical structures (Mengistu, Huizinga, Mouret, & Clune, 2016). The hierarchical segregation of neural networks into distributed neighbourhoods has also been found to stretch the parameter range for self-organized criticality by allowing subcritical and supercritical dynamics to co-exist simultaneously (Hilgetag & Hütt, 2014), which optimizes information processing and is therefore likely to be favoured by selection (Hesse & Gross, 2014). Maturational studies of neural networks throughout childhood and adolescence have further revealed that human cortical development mirrors phylogeny, progressing from the sensorimotor hierarchies found in all mammals through to recent association areas shared by humans and other primates (Gu et al., 2015). The phylogeny of the brain is also reflected across nested levels of neurophysiological organisation, ranging from genes and transcription factors through to synaptic epigenesis and the long-range connectivity that is thought to underpin

consciousness (Changeux, 2017). It is unsurprising, then, that a hierarchical neural structure has been found in every mammal examined to date (Finlay & Uchiyama, 2015).

The HMM rests on two assumptions. First, it subsumes the FEP as a biologically plausible EST of the evolution, development, form, and function of the brain. The FEP suggests that instead of just encoding a model of the world, the brain itself manifests a (hierarchical) phenotypic transcription of causal structure in its environment (i.e., an embodied statistical or generative model) that is optimized by evolution, development, and learning. Here, selection is simply nature's way of performing Bayesian model optimisation by minimising the (variational) free energy of human phenotypes across different (Tinbergian) timescales (Figure 4, panel C). Emphasis is placed on adaptive priors, which are (epi)genetically-specified expectations that have been shaped by selection to guide action-perception cycles toward adaptive or unsurprising states (Friston, 2010).

The second assumption follows EST in psychology by recognising that the brain is an emergent sub-system produced by a hierarchy of causal mechanisms that act on the brain-body-environment system over time (i.e., adaptation, phylogeny, ontogeny, and mechanism; Badcock, 2012; Kirchhoff, 2015, 2016). This causal hierarchy is recapitulated by major paradigms in psychology—which focus differentially on Tinbergen's (1963) four questions (Figure 4, panel C)—and calls for sophisticated, multiscale hypotheses that synthesise the FEP with different research programs in psychology and anthropology to explain both why a particular trait is adaptive, along with how it emerges from intergenerational, developmental, and real-time processes (Badcock, 2012; Badcock et al., 2017). The FEP and EST in psychology can therefore be regarded as different sides of the same coin—the FEP supplies a fully generalizable mechanistic theory of neural structure and function that applies to species generally, while an evolutionary systems approach to psychology builds upon the FEP by providing a substantive, evolutionary framework that is capable of explaining how the FEP manifests in Homo sapiens (Badcock, Ploeger, & Allen, 2016).

4.2. Translating variational neuroethology into research heuristics

The HMM is not just a second-order theory derived from variational neuroethology; it can also be translated into a multidisciplinary research heuristic that promotes a tripartite approach to scientific inquiry (Badcock et al., 2017). First, it calls for integrative, multilevel hypotheses in psychology that address all four of Tinbergen's questions by synthesising research across evolutionary psychology, biological and cognitive anthropology; evolutionary developmental biology and psychology; developmental psychology; and psychological sub-disciplines such

as cognitive, social, biological and personality psychology (Figure 4, panel C). Second, it requires an analysis of how the phenomenon of interest can be explained in terms of the FEP. Finally, it calls for empirical research into the ways in which the phenomenon is neurophysiologically instantiated in a dynamic, hierarchical manner and manifests behaviourally via active inference (Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2016). Although this synergistic approach to developing hypotheses is undoubtedly complex and remains in its infancy, it has already furnished important new insights into depression (Badcock et al., 2017), and yields distinct implications for neuroscientists and psychologists alike, while creating new avenues for collaborative research. This variational neuroethology of human systems has the major methodological advantage of conferring bona fide predictive power to the biological, cognitive, and social sciences. We can use simulation studies to predict which solutions will be adopted to respond adaptively to a given ecological problem, and then compare these computational models with empirical data.

For neuroscientists, the HMM lends itself to multimethod approaches that explore how detectable regularities in measurements of the brain can be explained by the psychological factors responsible for different patterns of neural activity in different contexts (Anderson, 2014). One way to isolate these factors is to use large databases of task-based fMRI activation studies to characterize the functional fingerprints of specific neural regions across different task demands (Anderson, 2014), while another is to use such datasets or combine functional activation studies with neuropsychological lesion-deficit models to derive cognitive ontologies that systematically map relationships between specific cognitive functions and hierarchical neural dynamics (Poldrack, 2010; Price & Friston, 2005). More broadly, the HMM also encourages the uptake of traditional methods in psychology—such as observational data collection and interviews—to explore the intersections between mind, brain, and behaviour (Badcock et al., 2017). Approaches in developmental psychology can be used to explore the dynamic ways in which human development differentiates (error-minimising) neurocognitive and biobehavioural patterns between individuals (Badcock et al., 2017), while comparative, cross-cultural, computational and dynamical approaches in evolutionary psychology and cognitive anthropology can elucidate the (epi)genetic mechanisms that underlie our species-typical adaptive priors (Badcock et al., 2016; Henrich, 2015; Ramstead et al., 2016; Tomasello, 2014). Finally, dynamical methods such as computer simulations and computational models enable us to directly examine how such levels of causation interact (Chiel & Beer, 1997; Frankenhuis, Panchanathan, & Clark Barrett, 2013; Friston et al., 2014), allowing neuroscientists to discover how the phenomena highlighted by psychologists reflect adaptive

free-energy minimisation under different (evolutionary, developmental and ecobiopsychosocial) conditions. The outcomes of such analyses might then be confirmed through experimental research, potentiating a fruitful dialectic between computational analyses and real-world observations.

On the other hand, the FEP offers a neurobiologically plausible EST of the mind and biobehaviour to psychologists. It has already been fruitfully applied to a wide range of psychological phenomena (Clark, 2015; Hohwy, 2014), extending from emotion (e.g., Joffily & Coricelli, 2013), anxiety (e.g., Hirsh, Mar, & Peterson, 2012) and depression (e.g., Badcock et al., 2017), through to autism (e.g., Van de Cruys et al., 2014), illusions (e.g., Hohwy, Roepstorff, & Friston, 2008) and psychosis (e.g., Corlett, Frith, & Fletcher, 2009). The FEP also lends itself to methods that are highly familiar to psychologists, such as the P300; a psychophysiological measure that captures temporal fluctuations in surprise (Nieuwenhuis, Aston-Jones, & Cohen, 2005). Finally, since it can be equally applied across all four of Tinbergen's research questions, the FEP casts new light on the reciprocal interplay of ultimate and proximate processes; stands to benefit both evolutionary and developmental psychologists; and proffers a common, transdisciplinary language to unite psychology's sub-disciplines (Badcock, 2012; Badcock et al., 2017).

However, if human phenotypes can only be understood by analysing the ecology from which they emerge, how might the HMM incorporate sociocultural influences—arguably our most influential selection pressure to date (Byrne & Whiten, 1988; Henrich, 2015; Tomasello, 2014)? A promising way to address this question is to incorporate recent work on cultural affordances. According to this perspective, cultural ensembles minimise free energy by enculturing their members so that they share common sets of precision-weighting priors (Ramstead et al., 2016). Human beings—with our specific forms of neural organisation, phenotypes, evolved behavioural tendencies and sociocultural patterns—minimise more free energy across spatial and temporal scales than any other species. Arguably, this is because we have crossed the 'evolutionary Rubicon': our survival depends on our ability to access and leverage cultural information and immersively participate in culturally adapted practices (Henrich, 2015). This is just another coordinated set of nested spatial and temporal scales in the greater Markov blanket of Homo sapiens. Another scale to consider—which we share with other animals like beavers and bees—is the free energy minimisation accomplished by designer environments (Clark, 2015). Although the ways in which the HMM might be extended to incorporate cultural affordances remains an open question, a promising avenue would be to explore approaches in scientometrics. Using science as the subject of its inquiry, this discipline

incorporates a wealth of models and quantitative techniques (e.g., citation and text analyses) that can be used to analyse how general selection and self-organisation act upon theorizing and research to reduce uncertainty about the environment (Leydesdorff, 1995, 2001). This affords a promising means to explore how the sociocultural generative models instantiated by different disciplines optimise (scientific) model selection by minimising free-energy (i.e., scientific prediction errors) over time (Badcock, 2012; Clark, 2013).

Ultimately, the HMM and theory of cultural affordances both rely on the idea that the Markov blanket of Homo sapiens possesses hierarchically nested temporal and spatial dimensions. These models both appeal to a scientific meta-theory (i.e., variational neuroethology) that defines Homo sapiens as each individual (phenotype) throughout the course of evolution, along with every individual that has either existed in the past or occupies the present (i.e., our species)—who may, in turn, alter the course of our evolution and our characteristically adapted niches. This is just another way of saying that the Markov blanket constituted by a single phenotype is dynamically nested within the global Markov blanket of our species (extending across all of Tinbergen's temporal scales). Conveniently, this approach also allows researchers to navigate the recursive complexities of the spatial axis (from biological macromolecules all the way up). By placing our Markov blanket around Homo sapiens, we necessarily encapsulate all of the dynamic, lower-level processes responsible for producing every phenotype, while imposing a clear upper limit on the complex adaptive system under scrutiny. Although the human Markov blanket is nested within the broader dynamics of other global Markov blankets that extend out into the universe, these lie beyond the limits of the system that this ecobiopsychosocial framework endeavours to explain.

5. Concluding remarks

The purpose of this article was to answer Schrödinger's question – 'what is life?' – by presenting a hierarchical multiscale free energy formulation—called variational neuroethology—that offers the sciences of life, mind, behaviour and society with a principled, computationally tractable guide to discovery. Arguably, the FEP affords a unifying perspective that explains the dynamics of living systems across spatial and temporal scales. The ontology for biological systems discussed here entails a multidimensional phase space populated by events (i.e., living systems as adaptive, self-organising patterns) that depend on the temporal dynamics of free energy minimisation. These events can be described as spatially and temporally nested Markov blankets, the dynamics of which are summarised by the Lagrangian of the FEP. Synthesising this framework with Tinbergen's four questions allows us to frame

our scientific investigations with clearly defined temporal and spatial bounds, depending on the specific questions we wish to answer. We have described an empirically supported exemplar of this approach in the cognitive and behavioural sciences—the HMM—and shown how this theory yields unique and promising avenues for interdisciplinary research. The challenge, of course, lies in translating theory into productive scientific practice, and in testing the current limits of the FEP by applying it to species without a brain, like E. coli, fungi and flora.

Supplementary materials:

Supplementary Information Box 1. Dynamical systems theory.

Evolutionary systems theory (EST), the ambient meta-theory that frames the free energy formulation, subsumes dynamical systems theory (DST). DST is a mathematical formulation of systems dynamics. Living systems have been productively modelled using the resources of DST, and so a few of the central concepts borrowed from DST by EST should be explicated. The core feature of dynamicist approaches is their emphasis on dynamics that unfold over time. The free energy formulation shares its dynamicist commitments with closely related approaches in cognitive neuroscience and psychology: the ecological and enactive approaches to cognition and behaviour. Indeed, the free energy formulation has been proposed as a computationally-tractable, mathematical formulation of enactivism from first principles (Allen & Friston, 2016; Bruineberg & Rietveld, 2014; Friston, 2013; Kirchhoff, 2015, 2016). For example, all the process theories that follow from the FEP, from predictive coding through to the belief propagation in decision-making, rest explicitly on formulating neuronal dynamics as a gradient descent on variational free energy (Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2017).

Dynamical systems are mathematical models that are used to represent physical systems with temporally extended dynamics. These dynamics are expressed as systems of (ordinary, stochastic or random) differential equations that describe trajectories or paths through phase space. They are defined over states of the system and their flow depends upon the current value of the system states. The analytic intractability of some of these systems of equations led to the development of qualitative methods to study dynamical systems. The strategy here is to qualitatively describe the dynamical evolution of a system by describing the abstract space of all of its possible states. A '*phase space*' is an abstract representation of all the possible states of a system. It is an *n*-dimensional (usually metric) space, where each dimension corresponds to a state variable of the system (e.g., position, velocity, etc.). A point in this *n*-dimensional phase space is an *n*-tuple that can be interpreted as assigning a value to each variable along each dimension. Because a point in the phase space of a dynamical system specifies a value for every variable of the system, any given point in this space represents a complete description of the system at a given instant.

How does this help us analyse the dynamics of a system? In dynamical systems theory, time is represented not as a separate dimension, but as a *dynamical trajectory* or sequence of states through the phase space. This trajectory is determined by the *topology* of the phase space,

which translates the constraints to which the system is subject. That is, the topology of the phase space translates the system of differential equations that govern the dynamics of the system being studied, usually in terms of attracting sets, manifolds or orbits (Arnold, 2003; Beer, 1995; Crauel & Flandoli, 1994; Freeman, 1994).

DST has especially been employed to understand *self-organisation*. Self-organisation is ubiquitous is nature: weather patterns, like rainstorms and lightning, emerge spontaneously; fluids crystallize to form structured lattices; bubbles arise in sea foam, and pop. This occurs because the spontaneous emergence of self-organized dynamics increases the efficiency of energy gradient dissipation and entropy production within that system (i.e., it increases its internal order). Self-organised dynamics emerge around an energy gradient, and optimise the flow of energy in the system about that energy gradient, until the system succumbs to decay (England, 2013). DST provides qualitative and computationally tractable formalisms to model such self-organising dynamics.

Supplementary Information Box 2. Variational inference and the free energy formulation.

Organisms do not have access to the 'true' probabilistic contingencies that describe the entire organism-niche system, that is, the actual relations of dependencies between environmental states and states of the organism. After all, the biotic system itself is 'hidden', as it were, behind a Markov blanket, which endows it with statistical independence from random fluctuations and other influences from the 'outside'. However, it does have access to quantities that define the variational free energy, and it can leverage the gradients defined by the free energy landscape to resists entropic erosion, through the process of 'active inference' (Friston, Kilner, & Harrison, 2006a; Friston, Daunizeau, Kilner, & Kiebel, 2010). Here, the organism's action-perception cycles can be seen as self-evidencing (Hohwy, 2016); that is, as producing evidence that allows it to infer its own existence.

Self-evidencing and variational inference

Formally, we can model this behaviour using the *variational* methods that were first developed by Feynman (1972), and are now widely used in statistical mechanics and machine learning (Dauwels, 2007; Hinton & Zemel, 1993; MacKay, 1995; Wallace & Dowe, 1999). In the context of the free energy formulation, the internal states of the organism (i.e., internal states of the Markov blanket) are formally described as encoding a 'variational' density, which comes to approximate the 'true' (posterior) probability density that is embodied by the organism-niche system, namely, through gradient descent on the free energy landscape. This warrants further explication.

Because living systems are inferentially secluded behind their Markov blanket, the causes of their sensory states are represented using surrogate or fictive variables, η, which represent the system's 'best guess' as to the cause of its input. As such, on the free energy formulation, the internal states of the organism encode or embody a 'variational density', which represents the organism's 'best guess' as to the causes of its sensations through cycles of free energy minimisation (a.k.a. active inference).

The ensuing formalism draws on approximate Bayesian inference (e.g., variational approaches in machine learning). When the computation of a posterior probability scheme becomes intractable in Bayesian inference (e.g., due to high dimensional and nonlinear generative models), a common strategy is to approximate the 'true' posterior density over the model parameters with a simpler variational density, whose sufficient statistics can be

optimised easily. Usually, this involves something called a *mean field approximation* in which a high dimensional posterior density is approximated with a product of marginal densities. These marginal densities then minimise variational free energy by passing messages (sufficient statistics) to each other. When this variational message passing is cast as a gradient descent on variational free energy, we obtain a DST version of approximate (variational) inference.

In the present (general) setting, the variational density is defined as a probability density, q, that is encoded or parameterized by internal states $\mu(t)$ of the system of interest, which is itself bounded by a Markov blanket and subject to free energy minimising dynamics. Thus, the internal states of the system induce a variational density over external states $\eta(t)$. This is a consequence of the formulation of the free energy as *surprise* plus *divergence*. This means that when the variational free energy functional is minimised, the divergence disappears and the variational density becomes the posterior density. At the same time, the variational free energy approximates surprise or (negative) log Bayesian evidence.

Active inference and variational free energy

So, although the organism cannot access the true posterior density (or surprise), it can nonetheless evaluate the variational free energy, because this quantity is a function of two quantities which it can access: the variational density that it encodes (which is parameterised by its internal states) and the sensations or sensory states of the Markov blanket that are contingent upon action. This brings us to a crucial observation. The only way we can actually change surprise is by changing the sensory states through action. This means to minimise surprise through minimising variational free energy, we need to change our internal states to make the free energy a good proxy for surprise and then we need to act to change our sensory states to actually reduce surprise. If we do this for long enough, the expected surprise or entropy of our sensory exchanges will be minimised (i.e., model evidence will be maximised) and we will be locally ergodic and self-evidencing. This is 'active inference'.

The generative models that define free energy are probabilistic models of the eco-niche relation, the dynamical relation that couples the organism to its niche. The generative model is usually expressed as the joint probability of sensory states, $s(t)$, and their causes, $\eta(t)$, in the (external) environment. This joint probability is usually expressed in terms of a likelihood $p(s(t)|\eta(t))$ and prior beliefs $p(\eta(t))$. With this formalism in place, one can say the variational density – that is parameterised by internal states – is *encoded* or *embodied* by the internal states of the system. Conversely, we can say the generative model is *entailed* by the existence of a

system equipped with a Markov blanket. Equivalently, we can also say that the generative model is *enacted* by the entire organism-niche system—that is, the conditional dependencies described by the generative model are brought forth in a process that is accomplished by exchanges between the organism and econiche.

Ergodicity and active inference

By virtue of the existence of an attracting or characteristic set of states (the extended phenotype), the dynamics of the random dynamical system must be locally *ergodic*. Living systems must maintain their ergodicity in order to remain alive – and thereby engage in some form of *active inference*. On this view, ergodicity is not merely an assumption that underwrites any principle of self-organisation. It is a definition of the living, complex adaptive (biological) systems we want to characterise. In other words, any system that does not maintain its (local) ergodicity cannot, by definition, possess characteristics that can be measured. The ergodicity of systems that possess a random dynamical attractor allows us to associate the long-term average of self-information or surprisal with the entropy of the probability distribution of occupying different states. This means that a tendency to minimise surprisal (or, equivalently, to maximise model evidence) is also a tendency to minimise (internal) entropy and thereby resist the second law of thermodynamics. In the sense that entropy is formally equivalent to the time average of a log probability, it corresponds to a Hamiltonian action. This means that all we are saying is that living systems conform to Hamilton's principle of least action, where action is entropy. In fact, we are saying a little bit more than this, we are saying that living systems conform to Hamilton's principle of least action via active inference – and the implicit minimisation of variational free energy.

Supplementary Information Box 3. Nested Markov blankets.

The ontology that we propose for living systems relies on the notion of nested Markov blankets. Markov blankets are a statistical description of dependencies in coupled systems. A Markov blanket comprises a set of systems states that separate internal and external states of any given system. More specifically, the existence of a Markov blanket entails a conditional independence between internal and external states. In turn, the Markov blanket itself can be divided into sensory and active states. These are characterized by the following relations: internal states cannot influence sensory states, while external states cannot influence active states. With these conditional independencies in place, we now have a well-defined (statistical) separation between the internal and external states of any system. This is relevant for our purposes because it tells us about what does, and does not, constitute a system.

To be equipped with a Markov blanket means the internal states can exhibit a selective openness to the outside, but only in a conditional sense. In the terms of dynamic systems theory, Markov blankets allow for random dynamical systems that are open to the external (environmental) states yet retain their own (statistical) integrity (Friston, 2013). They are open precisely to the extent that they exist far from thermodynamic equilibrium, and are permeable to fluctuations that originate from the environment. Open systems are open to energy and information exchange with their environment. However, internal systems are enclosed, because they are segregated (i.e., statistically independent) from external perturbations, and as such, are statistically insulated from environmental dynamics. These perturbations could lead to an altered topology (or attractor landscape) and an ensuing loss of integrity (phenotype, ergodicity). Clearly, then, for the living system to persist, its ergodicity needs to be maintained. This is evidenced by the persistence of the Markov blanket and its maintenance through active inference (Friston et al., 2016).

Now consider an ensemble of Markov blankets that exchange with each other. Recall that free energy is a functional of the beliefs entailed by internal states. However, the states of the Markov blanket will, in the ensemble, depend upon other Markov blankets. This means that the free energy minimum (when pooled over an ensemble of Markov blankets) necessitates a collective inference, in which no individual constituent of the ensemble surprises another. This coherence – perhaps mediated by generalised synchrony across Markov blankets – follows naturally from the free energy formalism (see Friston et al., 2015, for an example based upon cellular interactions in morphogenesis). Put simply, if we all have a common agenda and each play our role, fulfilling each other's expectations, we can collectively minimise surprise and maintain our ensemble of Markov blanket – and implicitly a Markov blanket of Markov

blankets. This provides the necessary milieu for each individual to thrive in a surprise-minimising, uncertainty-reducing sense; which, in turn, enables their brains to minimise free energy at the level of neurophysiology—all the way down to macromolecules and quantum physics.

This hierarchal composition of Markov blankets of Markov blankets follows naturally from the existence of a Markov blanket – the existence of Markov blankets around the system of interest is mandated by the existence of any system that can be distinguished from its external milieu. A crucial aspect of the hierarchical organisation of nested Markov blankets is that, at each level, the internal states are dropped from the game, as it were. In other words, the only states of interest—as we are lifted from one level to the next—are the Markov blankets that contain all of the necessary information for interactions at each level. Here, the dynamics of internal states are absorbed into the random fluctuations that are countered by gradient building, surprise minimising, self-evidencing flows. In other words, random fluctuations at one level are the inferential dynamics of internal states at the level below. The idea put forward in this paper is that as we ascend hierarchical scales, the fast fluctuations of internal states that are necessary to preserve the integrity of Markov blankets become fast random fluctuations that are averaged away at the scale above.

The hierarchical interdependencies provide a context for phenomena at each scale, which implies that the preservation of ergodic Markov blankets, at a lower scale, are necessary for the ongoing conditional independencies that form the basis of ergodic Markov blankets at a higher scale, and vice-versa. Thus, every level depends upon every other level: the central claim is that at every level, the same variational, surprise-reducing dynamics must be in play to provide Markov blankets for the next level. We call on this central observation to frame important relationships in hierarchical selection and self-organisation.

Supplementary Information Box 4. A gauge-theoretical free energy formulation and variational neuroethology

The 'intentionality' or 'aboutness' of living systems—that is, the directedness of the organism towards a meaningful world of significance and valence—emerges as a natural consequence of embedded adaptive systems that satisfy the constraints of the free energy formulation. For a living thing to be intentional just means that it entails a generative model (that necessarily includes prior beliefs about the way it should behave). To entail a generative model means that the organism vicariously enacts or brings forth the conditions that define it as a dynamically coupled, complex adaptive system. Put simply: active systems are alive if, and only if, there active inference entails a generative model. This makes the generative model of central importance to the free energy formulation, since it defines the form of life that an organism is seen to enact.

However, our modelling strategies need not be constrained by generative models. The nested Markov blanket formalism allows us to define a formal ontology of living systems independent of specific generative models. How? Recall that Markov blankets can be nested within Markov blankets, over spatial and temporal scales. This suggests that there has to be an internal scale-free or scale-invariant consistency: the free energy associated with a global Markov blanket (e.g., a species) must conform to the same principles as all of its constituent Markov blankets (e.g., phenotypes). This consistency must apply recursively, all the way down to the level of biological macromolecules.

The free energy principle has recently been reformulated using the resources of gauge theory (Sengupta et al., 2016). Gauge theory is a family of mathematical models that have been productively used in nearly all areas of physics, from electromagnetism to relativity and quantum mechanics. Gauge theories define a kind of dynamics, based on a Lagrangian, which preserves a symmetry or invariant. This symmetry is broken by local forces. In turn, symmetry breaking recruits a gauge field, which compensates for the local perturbations, restoring the system symmetry.

The free energy functional has been proposed as the Lagrangian of a gauge theory for neural and biological systems (Sengupta et al., 2016). Crucially, this means that it is possible to mathematically model invariant free energy minimisation dynamics across temporal and spatial scales. The gauge theoretical Lagrangian, after all, is constructed to be scale invariant. This kind of modelling is less constrained by specific generative models, because it is interested in the resolution of local dynamics through the effects of gauge fields (which, here, is interpreted as free energy minimisation via active inference). Thus, over any period of time,

the scale-invariant free energy Lagrangian is minimised across spatial scales, recruiting different local gauge dynamics as a function of the different local forces. At each scale, the local dynamics, determined by local Markov blanket features, perturb the Lagrangian. These perturbations are then compensated by gauge fields, which in the free energy formulation are achieved through active inference (perception, action, evolution, niche construction, and so forth).

This means that we can draw broad equivalence classes between different forms of free energy minimisation and thereby connect behavioural patterns that, while isomorphic from the standpoint of free energy minimisation, are mechanistically implemented or realised in different ways. Thus, the gauge-theoretical formulation allows us to define classes of equivalent internal states, patterns of action, and generative models (pitched at different temporal and spatial scales) that minimise free energy in systematically or dynamically equivalent, but mechanistically heterogeneous, physical systems. The gauge theoretical framework suggests that different classes of generative models are able to guide the same (or analogous) dynamics, which points towards the possibility of developing a computational neuroethology based on the free energy formulation—which we have called 'variational neuroethology'.

Living systems that conform to the principles described here will appear to minimise their free energy. However, this minimisation of free energy is not an additional force of nature or élan vital. Rather, it is a fictive force, which is ultimately due to the geometry of information. The probabilistic landscape over which state transitions for living systems are defined has a nontrivial curvature, since they are able to perform gradient descent. For example, in the context of predictive coding in the brain, this curvature can be interpreted as attention (Sengupta et al., 2016). Living systems are organized such that their (physical) dynamics are also, at the same time, coextensive with trajectories over information landscapes. In other words, the FEP tells us that for an organism to follow Hamilton's path of least action simply means that it must instantiate (embody and enact) a statistical model of its relation to its niche. 'Inference', then, is just the process whereby mutual information between organism and niche increases through generalized synchrony; it is not 'inference' in some strong propositional sense, but rather, a precise mathematical one (i.e., approximate Bayesian inference).

References

Allen, M., & Friston, K. J. (2016). From cognitivism to autopoiesis: towards a computational framework for the embodied mind. *Synthese*, 1-24.

Anderson, M. L. (2014). *After phrenology : neural reuse and the interactive brain*. Cambridge, MA; London, England: The MIT Press.

Anderson, M. L., & Finlay, B. L. (2014). Allocating structure to function: the strong links between neuroplasticity and natural selection. *Frontiers in human neuroscience, 7*.

Ao, P. (2003). Stochastic force defined evolution in dynamical systems. *arXiv preprint physics/0302081*.

Ao, P. (2005). Laws in Darwinian evolutionary theory. *Physics of life Reviews, 2*(2), 117-156.

Ao, P. (2008). Emerging of stochastic dynamical equalities and steady state thermodynamics from Darwinian dynamics. *Communications in theoretical physics, 49*(5).

Ao, P. (2009). Global view of bionetwork dynamics: adaptive landscape. *Journal of Genetics and Genomics, 36*(2), 63-73.

Ao, P., Chen, T.-Q., & Shi, J.-H. (2013). Dynamical decomposition of markov processes without detailed balance. *Chinese Physics Letters, 30*(7), 070201.

Arnold, L. (2003). *Random Dynamical Systems (Springer Monographs in Mathematics)*. Berlin: Springer-Verlag.

Badcock, P. B. (2012). Evolutionary systems theory: a unifying meta-theory of psychological science. *Review of General Psychology, 16*(1), 10-23.

Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017). The depressed brain: an evolutionary systems theory. *Trends in Cognitive Sciences, 21*, 182-194.

Badcock, P. B., Ploeger, A., & Allen, N. B. (2016). After phrenology: time for a paradigm shift in cognitive science. *Behavioral and Brain Sciences, 39*. doi:doi.org/10.1017/S0140525X15001557

Bak, P., Tang, C., & Wiesenfeld, K. (1988). Self-organized criticality. *Physical review A, 38*(1), 364-374.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron, 76*(4), 695-711.

Beer, R. D. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence, 72*(1), 173-215.

Blackmore, S. (2000). *The Meme Machine*. Oxford: Oxford University Press.

Bruineberg, J., Kiverstein, J., & Rietveld, E. (2016). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese*, 1-28. doi:doi:10.1007/s11229-016-1239-1

Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in human neuroscience, 8*. doi:doi.org/10.3389/fnhum.2014.00599

Buckner, R. L., & Krienen, F. M. (2013). The evolution of distributed association networks in the human brain. *Trends in Cognitive Sciences, 17*(12), 648-665.

Byrne, R. W., & Whiten, A. (Eds.). (1988). *Machiavellian Intelligence: Social Expertise and Evolution of Intellect in Monkeys, Apes and Humans*. Oxford: Oxford University Press.

Campbell, J. O. (2016). Universal Darwinism as a process of Bayesian inference. *Frontiers in Systems Neuroscience, 10*.

Caporael, L. R. (2001). Evolutionary psychology: Toward a unifying theory and a hybrid science. *Annual review of psychology, 52*(1), 607-628.

Changeux, J.-P. (2017). Climbing brain levels of organisation from genes to consciousness. *Trends in Cognitive Sciences, 21*(3), 168-181.

Chiel, H. J., & Beer, R. D. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in neurosciences, 20*(12), 553-557.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181-204.

Clark, A. (2015). *Surfing uncertainty: prediction, action, and the embodied mind*. Oxford: Oxford University Press.

Clark, A. (2017). How to knit your own Markov blanket. In T. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing*. Frankfurt am Main: MIND Group.

Conant, R. C., & Ross Ashby, W. (1970). Every good regulator of a system must be a model of that system. *International journal of systems science, 1*(2), 89-97.

Corlett, P. R., Frith, C. D., & Fletcher, P. C. (2009). From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology-Berlin, 206*(4), 515-530.

Crauel, H., & Flandoli, F. (1994). Attractors for random dynamical systems. *Probab Theory Relat Fields, 100*, 365-393.

Cziko, G. (1995). *Without miracles: universal selection theory and the second Darwinian revolution*. Cambridge, MA: MIT press.

Dauwels, J. (2007, 24-29 June 2007). *On Variational Message Passing on Factor Graphs.* Paper presented at the 2007 IEEE International Symposium on Information Theory.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The helmholtz machine. *Neural computation, 7*(5), 889-904.

Depew, D. J., & Weber, B. H. (1995). *Darwinism evolving: systems dynamics and the genealogy of natural selection*. Cambridge, MA: MIT Press.

Di Paolo, E. A., & Thompson, E. (2014). The enactive approach. In L. E. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 68-78). London: Routledge.

Dorogovtsev, S. N., & Mendes, J. F. (2002). Evolution of networks. *Advances in physics, 51*(4), 1079-1187.

Eigen, M., & Schuster, P. (1979). *The hypercycle: a principle of natural self-organisation*. Berlin: Springer-Verlag.

England, J. L. (2013). Statistical physics of self-replication. *J Chem Phys, 139*(12), 121923. doi:10.1063/1.4818538

Feynman, R. (1972). *Statistical mechanics: a set of lectures*. Reading, MA: Benjamin/Cummings Publishing.

Finlay, B. L., & Uchiyama, R. (2015). Developmental mechanisms channeling cortical evolution. *Trends in neurosciences, 38*(2), 69-76.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection* Oxford Clarendon Press

Frank, S. A. (2012). Natural selection. V. How to read the fundamental equations of evolutionary change in terms of information theory. *Journal of evolutionary biology, 25*(12), 2377-2396.

Frankenhuis, W. E., Panchanathan, K., & Clark Barrett, H. (2013). Bridging developmental systems theory and evolutionary psychology using dynamic optimization. *Developmental Science, 16*(4), 584-598.

Freeman, W. J. (1994). Characterization of state transitions in spatially distributed, chaotic, nonlinear, dynamical systems in cerebral cortex. *Integr Physiol Behav Sci., 29*(3), 294-306.

Friston, K. (2011). Embodied inference: or ‘‘I think therefore I am, if I am what I think’’. In W. Tschacher & C. Bergomi (Eds.), *The implications of embodiment: Cognition and communication* (pp. 89-125). Exeter, UK: Imprint Academic.

Friston, K. (2012). A free energy principle for biological systems. *Entropy, 14*(11), 2100-2121.

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

Friston, K., & Ao, P. (2011). Free energy, value, and attractors. *Computational and mathematical methods in medicine, 2012*.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active Inference: A Process Theory. *Neural Comput, 29*(1), 1-49. doi:10.1162/NECO_a_00912

Friston, K., Kilner, J., & Harrison, L. (2006a). A free energy principle for the brain. *J Physiol Paris, 100*(1-3), 70-87. doi:10.1016/j.jphysparis.2006.10.001

Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360*(1456), 815-836.

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127-138.

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biol Cybern., 102*(3), 227-260.

Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2016). Active inference: a process theory. *Neural computation, 29*, 1-49.

Friston, K. J., & Frith, C. D. (2015). Active inference, communication and hermeneutics. *cortex, 68*, 129-143.

Friston, K. J., Kilner, J., & Harrison, L. (2006b). A free energy principle for the brain. *Journal of Physiology-Paris, 100*(1), 70-87.

Friston, K. J., Levin, M., Sengupta, B., & Pezzulo, G. (2015). Knowing one's place: a free-energy approach to pattern regulation. *Journal of the Royal Society Interface, 12*(105).

Friston, K. J., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. *The Lancet Psychiatry, 1*(2), 148-158.

Gallagher, S., & Allen, M. (2016). Active inference, enactivism and the hermeneutics of social cognition. *Synthese*, 1-22.

Gu, S., Satterthwaite, T. D., Medaglia, J. D., Yang, M., Gur, R. E., Gur, R. C., & Bassett, D. S. (2015). Emergence of system roles in normative neurodevelopment. *Proceedings of the National Academy of Sciences, 112*(44), 13681-13686.

Haken, H. (1983). *Synergetics: An introduction. Non-equilibrium phase transition and self-selforganisation in physics, chemistry and biology*. Berlin: Springer-Verlag.

Haken, H. (1996). *Principles of brain functioning: a synergetic approach to brain activity, behaviour and cognition*. Berlin: Springer-Verlag.

Harper, M. (2011). Escort evolutionary game theory. *Physica D: Nonlinear Phenomena, 240*(18), 1411-1415.

Henrich, J. (2015). *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton, NJ: Princeton University Press.

Henriques, G. (2011). *A new unified theory of psychology*. New York, NY: Springer

Hesse, J., & Gross, T. (2014). Self-organized criticality as a fundamental property of neural systems. *Frontiers in Systems Neuroscience, 8*(166).

Hilgetag, C. C., & Hütt, M.-T. (2014). Hierarchical modular brain connectivity is a stretch for criticality. *Trends in Cognitive Sciences, 18*(3), 114-115.

Hinton, G. E., & Zemel, R. S. (1993). *Autoencoders, minimum description length and Helmholtz free energy*. Paper presented at the Proceedings of the 6th International Conference on Neural Information Processing Systems, Denver, Colorado.

Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: a framework for understanding uncertainty-related anxiety. *Psychological review, 119*(2), 304-320.

Hohwy, J. (2014). *The predictive mind*. Oxford: Oxford University Press.

Hohwy, J. (2016). The self-evidencing brain. *Noûs, 50*(2), 259-285.

Hohwy, J., Roepstorff, A., & Friston, K. J. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition, 108*(3), 687-701.

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters, 78*(14), 2690-2693.

Joffily, M., & Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLOS Computational Biology, 9*(6).

Jung, M., Hwang, J., & Tani, J. (2015). Self-organization of spatio-temporal hierarchy via learning of dynamic visual image patterns on action sequences. *PloS one, 10*(7).

Kaiser, M., Hilgetag, C. C., & Kötter, R. (2010). Hierarchy and dynamics of neural networks. *Frontiers in Neuroinformatics, 4*(112), 4-6.

Kauffman, S. A. (1993). *The origins of order: self-organization and selection in evolution*. Oxford: Oxford University Press.

Kelso, J. S. (1995). *Dynamic patterns: the self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLOS Computational Biology, 4*(11).

Kiebel, S. J., & Friston, K. J. (2011). Free energy and dendritic self-organization. *Frontiers in Systems Neuroscience, 5*.

Kirchhoff, M. (2015). Species of realization and the free energy principle. *Australasian Journal of Philosophy, 93*(4), 706-723.

Kirchhoff, M. (2016). Autopoiesis, free energy, and the life–mind continuity thesis. *Synthese*, 1-22. doi:doi:10.1007/s11229-016-1100-6

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in neurosciences, 27*(12), 712-719.

Levin, S. (1998). Ecosystems and the biosphere as complex adaptive systems. *Ecosystems, 1*(5), 431-436.

Levin, S. (2003). Complex adaptive systems: exploring the known, the unknown and the unknowable. *Bulletin of the American Mathematical Society, 40*(1), 3-19.

Leydesdorff, L. (1995). *The challenge of scientometrics: The development, measurement, and self-organization of scientific communications*. Leiden: DSWO Press.

Leydesdorff, L. (2001). *A sociological theory of communication: The self-organization of the knowledge-based society*. Parkland, FL: Universal Publishers.

MacKay, D. J. (1995). Free-energy minimisation algorithm for decoding and cryptoanalysis. *Electronics Letters, 31*, 445-447.

Mantegna, R. N., & Stanley, H. E. (1995). Scaling behaviour in the dynamics of an economic index. *Nature, 376*(6535), 46-49.

Martyushev, L., & Seleznev, V. (2006). Maximum entropy production principle in physics, chemistry and biology. *Physics reports, 426*(1), 1-45.

Mengistu, H., Huizinga, J., Mouret, J.-B., & Clune, J. (2016). The evolutionary origins of hierarchy. *PLOS Computational Biology, 12*(6).

Mesoudi, A., Whiten, A., & Laland, K. N. (2006). Towards a unified science of cultural evolution. *Behavioral and Brain Sciences, 29*(04), 329-347.

Miller, J. H., & Page, S. E. (2009). *Complex adaptive systems: an introduction to computational models of social life*. Princeton, NJ: Princeton University Press.

Nicolis, G., & Prigogine, I. (1977). *Self-organization in nonequilibrium systems*. New York, NY: John Wiley.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological bulletin, 131*(4), 510-532.

Park, H.-J., & Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science, 342*(6158).

Poldrack, R. A. (2010). Mapping mental function to brain structure: how can cognitive neuroimaging succeed? *Perspectives on Psychological Science, 5*(6), 753-761.

Price, C. J., & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology, 22*(3-4), 262-275.

Prigogine, I., & Stengers, I. (1984). *Order out of chaos: man's new dialogue with nature*. New York, NY: Bantam Books

Ramstead, M., Veissière, S., & Kirmayer, L. (2016). Cultural affordances: scaffolding local worlds through shared intentionality and regimes of attention. *Frontiers in Psychology, 7*.

Rao, R. P., & 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.

Sarkar, S. (Ed.) (1992). *The Founders of Evolutionary Genetics: A Centenary Reappraisal* (Vol. 142). Dordrecht / Boston / London: Kluwer Academic Publishers

Schrödinger, E. (1944). *What is life?* Cambridge: Cambridge University Press.

Seifert, U. (2012). Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics, 75*(12).

Sengupta, B., Tozzi, A., Cooray, G. K., Douglas, P. K., & Friston, K. J. (2016). Towards a neuronal gauge theory. *PLOS Biology, 14*(3).

Seth, A. K. (2014). The cybernetic brain: from interoceptive inference to sensorimotor contingencies. In T. Metzinger & J. M. Windt (Eds.), *Open MIND* (pp. 1-24). Frankfurt am Main: MIND Group.

Shipp, S. (2016). Neural elements for predictive coding. *Frontiers in Psychology, 7*.

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B: Biological Sciences, 216*(1205), 427-459.

Su, H., Wang, G., Yuan, R., Wang, J., Tang, Y., Ao, P., & Zhu, X. (2017). Decoding early myelopoiesis from dynamics of core endogenous network. *Science China Life Sciences, 60*(6), 627-646.

Tang, Y., Yuan, R., & Ao, P. (2014). Summing over trajectories of stochastic dynamics with multiplicative noise. *The Journal of chemical physics, 141*(4), 044125.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie, 20*, 410-433.

Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, MA: Harvard University Press.

Tomé, T. (2006). Entropy production in nonequilibrium systems described by a Fokker-Planck equation. *Brazilian journal of physics, 36*(4A), 1285-1289.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychological review, 121*(4), 649-675.

Wallace, C. S., & Dowe, D. L. (1999). Minimum Message Length and Kolmogorov Complexity. *The Computer Journal, 42*(4), 270-283. doi:10.1093/comjnl/42.4.270

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In D. F. Jones (Ed.), *Proceedings of the Sixth International Congress of Genetics* (Vol. 1). Ithaca, New York: Brooklyn Botanical Garden.

Yuan, R., Zhu, X., Wang, G., Li, S., & Ao, P. (2017). Cancer as robust intrinsic state shaped by evolution: a key issues review. *Reports on Progress in Physics, 80*(4), 042701.

8.3. Chapter 3: Variational ecology and the physics of sentient systems

Original publication details:

Ramstead, M. J. D., Constant, A., Badcock, P. B., & Friston, K. J. (2019). Variational ecology and the physics of sentient systems. *Physics of Life Reviews*.
Invited contribution to the Special Issue of *Physics of Life Reviews* on Physics of the Mind, guest edited by Félix Schoeller.
doi.org/10.1016/j.plrev.2018.12.002.

Authors:
Maxwell J. D. Ramstead[1,2*]†
Axel Constant[3]†
Paul B. Badcock[4,5,6]
Karl J. Friston[7]

Affiliations:
[1]Division of Social and Transcultural Psychiatry, Department of Psychiatry, McGill University, Montreal, QC, Canada, H3A 1A1.
[2]Department of Philosophy, McGill University, Montreal, QC, Canada, H3A 2T7.
[3]Amsterdam Brain and Cognition Center. The University of Amsterdam, Amsterdam, The Netherlands, 1098 XH.
[4]Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, Australia, 3010.
[5]Centre for Youth Mental Health, The University of Melbourne, Melbourne, Australia, 3052.
[6]Orygen, the National Centre of Excellence in Youth Mental Health, Melbourne, Australia, 3052.
[7]Wellcome Trust Centre for Neuroimaging, University College London, London, UK, WC1N 3BG.
*Correspondence to: maxwell.ramstead@mail.mcgill.ca
† Contributed equally

Abstract:

This paper addresses the challenges faced by multiscale formulations of the variational (free energy) approach to dynamics that obtain for large-scale ensembles. We review a framework for modelling complex adaptive control systems for multiscale free energy bounding organism-niche dynamics, thereby integrating the modelling strategies and heuristics of variational neuroethology with a broader perspective on the ecological nestedness of biotic systems. We extend the multiscale variational formulation beyond the action-perception loops of individual organisms by appealing to the variational approach to niche construction to explain the dynamics of coupled systems constituted by organisms and their ecological niche. We suggest that the statistical robustness of living systems is inherited, in part, from their eco-niches, as niches help coordinate dynamical patterns across larger spatiotemporal scales. We call this approach variational ecology. We argue that, when applied to cultural animals such as humans, variational ecology enables us to formulate not just a physics of individual minds, but also a physics of interacting minds across spatial and temporal scales – a physics of sentient systems that range from cells to societies.

Keywords:

Physics of the mind; Free energy principle; Evolutionary systems theory; Variational neuroethology; Variational ecology; Niche construction.

Highlights:

We extend the multiscale variational free energy approach to large-scale ensembles
We integrate multiscale modelling with a broad perspective on ecological nestedness
We argue that the statistical robustness of living systems is ecologically inherited
We propose variational ecology as a physics of sentient systems

Acknowledgements:

Work on this article was supported by the Wellcome Trust (K. Friston: Principal Research Fellowship; Ref: 088130/Z/09/Z) and the Social Sciences and Humanities Research Council of Canada (M. J. D. Ramstead). We thank Casper Hesp, Samuel Veissière, and Laurence Kirmayer for their insightful critical remarks on variational neuroethology in this journal, which prompted us to revisit our account.

In the previous chapter,

I proposed a novel approach to neuroethology, i.e., the study of adaptive behaviour and its control in multiscale living systems, called variational neuroethology (VNE). VNE synthesises Tinbergen's seminal approach to ethology (centred on the analysis of function, phylogeny, ontogeny, and mechanism) with a mathematical model of the information-theoretic constraints that must be true of any self-organizing cognitive system maintaining itself in nonequilibrium steady state; namely, active inference under the free-energy principle. An outstanding issue for VNE is the question of how to scale up the framework in a principled manner to the study of systems beyond the organismic boundary; that is, the question how to draw the boundaries of cognitive systems the dynamics of which unfold at scales beyond, and include, that of the organism. The aim of this chapter is to address this issue from first principles. I propose an extension of VNE to more general ecological dynamics, called variational ecology (VE). VE integrates VNE with a variational approach to ecological niche construction. VE thus provides an approach to cognitive ecology that accounts for how hierarchically structured cognitive systems – from cells to societies – construct, and align themselves with, their ecological niche.

1. Introduction

Recent decades have witnessed the emergence of a new project for a *physics of the mind*. This effort has leveraged the constructs, principles, and methods of theoretical and experimental physics to investigate and understand what we call sentience and the 'mind'. At the turn of the century, a new evolutionary systems theory of complex adaptive systems was proposed, made possible by the advent of computational modeling, variational methods of inference, and machine learning: the (variational) free energy principle (FEP) (Friston, 2013; Friston, 2010; Friston, Kilner, & Harrison, 2006).

According to the FEP, the physics of sentient systems follows from the statistical mechanics of life. This variational formulation stems from the observation that living systems, over time and on average, tend to revisit the same set of *attracting* or *characteristic states*. These can be cast as the characteristic *phenotypic states* (and *traits*) of the organism. The FEP explains the dynamics of *any random dynamical system* that appears to resist decay through *adaptive action* (Friston, 2013; Ramstead, Badcock, & Friston, 2018). Under the FEP, organisms engage the environment in a self-fulfilling prophecy of sorts; 'surfing' up probability gradients towards their most probable phenotypic states (Clark, 2015). The FEP was originally proposed in computational neuroscience to explain neural dynamics (Friston, 2005; Friston & Stephan, 2007), where it coheres broadly with predictive coding approaches (Dayan, Hinton, Neal, & Zemel, 1995; Knill & Pouget, 2004; Metzinger & Wiese, 2017; Rao & Ballard, 1999), and is widely recognized as a unifying theory of the function, structure, and dynamics of the brain (Clark, 2013, 2015; Hohwy, 2014; Huang, 2008). It has since been extended to explain the dynamics of biological systems within and beyond the brain, ranging from the cellular level (Kiebel & Friston, 2011) and action-perception loops (Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2016b), to psychiatry (Adams, Huys, & Roiser, 2016; Constant, Bervoets, Hens, & Van de Cruys, 2018; Corlett & Fletcher, 2014; Lawson, Rees, & Friston, 2014; Redish & Gordon, 2016; Van de Cruys et al., 2014), psychology (Apps & Tsakiris, 2014; Badcock, 2012; Badcock, Davey, Whittle, Allen, & Friston, 2017; Hohwy, 2014) and embodied cognitive science (Bruineberg, Kiverstein, & Rietveld, 2016), through to evolutionary dynamics (cf. Campbell, 2016; Friston, 2011; Friston & Ao, 2011; Friston & Stephan, 2007) and, recently, to intersubjective and sociocultural dynamics (Fabry, 2017; Friston & Frith, 2015; Kirmayer & Ramstead, 2017; Ramstead, Veissière, & Kirmayer, 2016).

The FEP has recently been leveraged to furnish a fully generalizable metatheory for adaptive behaviour in sentient systems across spatial and temporal scales, called *variational neuroethology* (VNE) (Ramstead et al., 2018). VNE synthesises the FEP with Tinbergen's

(Tinbergen, 1963) four levels of explanation in biology (i.e., adaptation, phylogeny, ontogeny, and mechanism) to provide a systematic guide for theorising and research in the sciences of life and mind. As a metatheory, VNE comprises two principal components: a multiscale theoretical ontology for living systems based on a recursively nested formulation of Markov blankets (MBs); and a multidisciplinary research heuristic in the biological and cognitive sciences (Carr, 1981). In their response to peer commentaries, Ramstead and colleagues (2018) address the promise and limitations of VNE as a heuristic for scientific inquiry, and review its scope as a research programme. This paper will complement this discussion by concentrating on criticisms of VNE as a *multiscale ontology*.

To date, VNE has only provided a *principled method* of analysing nested and mutually constraining biological systems and their complex adaptive dynamics across spatiotemporal scales (Carr, 1981; Ramstead et al., 2018). It has yet to offer a way to individuate systems at scales beyond that of the organism acting in its environment; e.g., large-scale ensembles such as societies or ecosystems that realise free energy bounding dynamics. This issue was cogently articulated by Bruineberg and Hesp (2018) and Kirmayer (2018) in their critique of VNE. They asked whether the MB formalism leveraged by VNE to individuate systems is adequate for modelling phenomena at scales beyond those of a single organism (e.g., sociocultural dynamics); since phenomena at these scales may be too transient or not sufficiently robust to license the MB formalism (which is defined in terms of conditional independencies in weakly mixing random dynamical systems). Could this reflect a fundamental distinction between the FEP as an explanatory principle for clearly bounded, ergodic, biological systems, and the greater, complementary forces at play that constrain complex adaptive systems in general (including groups of organisms and their ecological niches)? Should we restrict the FEP to the level of the organism, and then explore how this model connects meaningfully to other key concepts in evolutionary systems theory about the agent-niche relation?

The aim of this paper is to address the challenges faced by VNE with regard to dynamics that obtain for large-scale ensembles. Specifically, we review a framework for modelling *complex adaptive systems* for multiscale free energy bounding organism-niche dynamics, thereby integrating the modelling strategies and heuristics of VNE with a broader perspective on the ecological nestedness of biotic systems. We extend VNE beyond the action-perception loops of individual organisms (i.e., active inference, Friston, Daunizeau, & Kiebel, 2009; Friston, Rosch, Parr, Price, & Bowman, 2018) by appealing to the *variational approach to niche construction* (VANC) (Constant, Ramstead, Veissière, Campbell, & Friston, 2018) to explain the dynamics of coupled systems constituted by organisms and their ecological niche.

We suggest that the statistical robustness of living systems is inherited, in part, from their eco-niches, as niches help coordinate dynamical patterns across larger spatiotemporal scales. We call this approach *variational ecology* (VE), which subsumes VNE and the VANC[7]. When applied to cultural animals such as humans, VE has the important consequence of allowing us to formulate not just a physics of individual minds, but also a physics of *interacting minds* across spatial and temporal scales – a *physics of sentient systems* that range from cells to societies.

2. The variational (free energy) formulation

The free energy formulation appeals to a *statistical conception of life*. It rests on the fact that on average and over time, living systems endure as bounded, self-organising partitions of dynamical systems. In other words, organisms appear to counter dissipative environmental perturbations by resisting the weathering effects of entropic decay dictated by the fluctuation theorems that hold at nonequilibrium steady-state (Evans & Searles, 2002; Seifert, 2012). Technically, biological systems revisit the same set of characteristic states that constitute a *random dynamical attractor*. This attracting set means they have properties that can be measured (i.e., they are locally ergodic). Put another way, an organism possesses an attracting set of states that it tends to occupy with a much higher frequency than others. The FEP provides a formal description of how organisms resist entropic erosion and maintain themselves within their phenotypic bounds. More exactly, it describes the dynamics they must exhibit, if they possess characteristic or attracting states.

For an organism to exist as a bounded system means that it must be able to maintain itself as a whole. By definition, for a system to exist at all, it must evince a robust form of *conditional independence* with respect to external (non-systemic) states. The variational framework addresses the individuation of relevant systems by operationalising the notion of conditional independence using the *Markov blanket* (MB) *formalism*. For a set of states to be enshrouded by a MB means that the *dynamics* of that system induce a *statistical partition* of its states (Friston, 2013; Palacios, Razi, Parr, Kirchhoff, & Friston, 2017; Ramstead et al., 2018). A MB is the set of states that statistically isolates (insulates) *internal* (systemic) from *external* (non-systemic) states, such that changes in internal states are mediated by the states

[7] Note that we use the term ‘ecology’ non-technically, to capture the sorts of interactions among agents and their environment, leading to intentional behaviour, or ‘mind’. We do not appeal directly to the science of ecology, or to human ecology.

of the MB. The MB itself can be partitioned into *active* and *sensory* states, which are defined by the following relations: internal states do not influence sensory states, and external states do not influence active states.

Now, we should note that the terms 'active' and 'sensory' are potentially misleading. They are only meant to capture relations of *statistical dependence* between random variables. This will be crucial to our argument below, as things that we would not readily describe as literally acting or sensing in any meaningful sense can still be captured with this formalism, since it entails only a statistical enshrouding of systemic states from external ones, and the systematic statistical partition of the whole organism-niche system (Ramstead et al., 2018).

2.1. Active inference and generative models

To keep the MB in play is not a trivial matter. The FEP explains the emergence and maintenance of this existential boundary. Under the FEP, the Markov boundary is dynamically enabled by *adaptive action*. The idea behind the FEP is that the *active behaviour* of organisms maintains them in states of viable, adaptive coupling with their ecological niche. *Active inference* is the process whereby living systems act on the world – and update their internal states – so as to embody or encode the statistical structure of their local environment, leading to the *adaptive control of behaviour* (Friston et al., 2009; Friston, Daunizeau, Kilner, & Kiebel, 2010). In other words, over time and on average, biological systems come to fit their environment, via active optimization procedures; e.g., action, perception, and learning (Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2016a). From a developmental perspective, this can be cast as phenotypic accommodation and developmental plasticity (Stearns, 1989; West-Eberhard, 2003).

Crucially, this functionalist interpretation is underwritten by the gradient flows implied by the existence of a random dynamical attractor (and implicit ergodicity). This is fairly straightforward to show using either a path integral or Fokker Planck formulation of density dynamics; in which the flow of states must climb probability gradients to counter the dispersive effects of random fluctuations. When we put the MB in play, the same gradient flows reveal an (en)active and adaptive interpretation of active states, which are informed by, and only by, sensory and internal states. This follows because the flows of active and internal states are special, in the sense that their flow does not depend upon external states. These structured dependencies, which necessarily follow from the existence of random dynamical attractors with MBs, lead to active inference.

In active inference, *active* and *internal states* of the organism's *MB* can always be expressed as minimising or *bounding* a quantity called *variational free energy* (see Figure 1). We can cast variational free energy as the disattunement between the statistical structure of the ecological niche and that of the organism (i.e., its phenotype and behavioural dynamics) (Bruineberg et al., 2016; Bruineberg & Rietveld, 2014). Technically, variational free energy is a *proxy* for a quantity called surprise (a.k.a. *surprisal*), which reflects the improbability of finding an organism in some sensory state (technically, surprise is the self-information or negative log probability of sensory samples encountered by an agent) (Friston, 2010). The organism cannot evaluate this quantity directly. Instead, it (or rather, its behaviour or dynamics) bounds a *proxy quantity*, which is a variational bound on surprise, in the sense that surprise can never be greater than this bound. This is variational free energy and is exactly the same quantity used in machine learning and variational Bayes (often referred to as an *evidence bound*). While surprise depends only on states of the world, variational free energy depends on a Bayesian belief or probability density that is encoded by its internal states. Through active inference, the probability densities entailed by organismic (internal) states tune themselves to effectively infer the process by which sensory states were generated; i.e., the dynamics of external, unobserved states that cause fluctuations in sensory states and are hidden behind the MB.

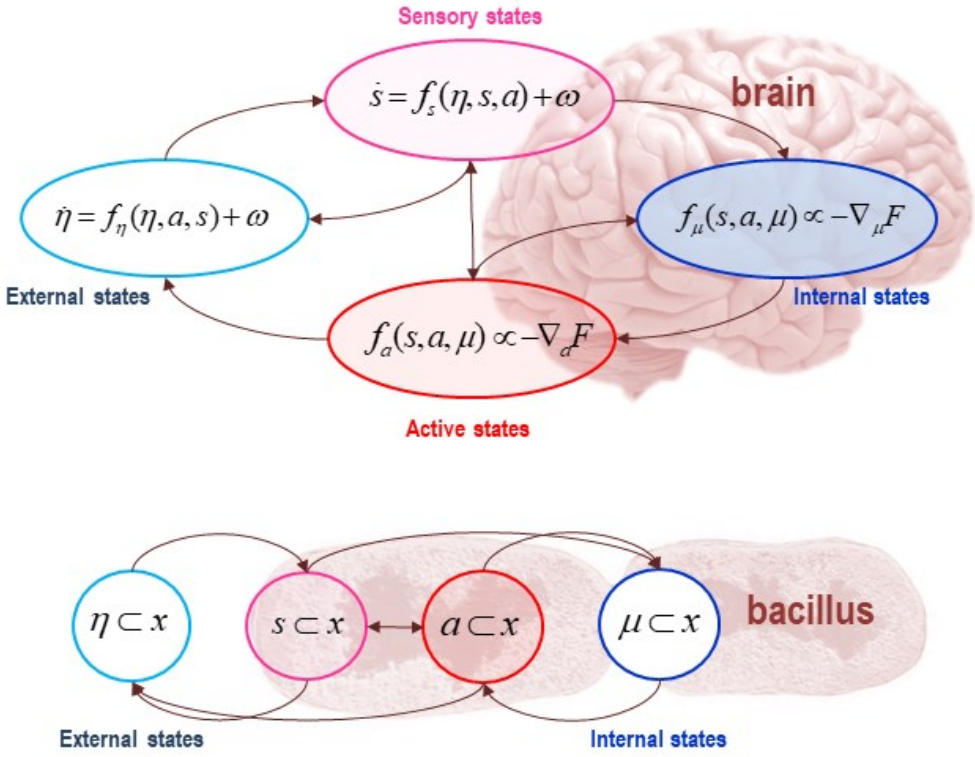


Fig. 1. *The Markov blanket*. These schematics illustrate the partition of states into internal (***μ***) and hidden or external states (***η***) that are separated by a MB – comprising sensory (***s***) and active states (***a***). The upper panel shows this partition as it would be applied to action and perception in the brain; where active and internal states minimise a free energy functional of sensory states. The ensuing self-organisation of internal states then corresponds to perception, while action couples brain states back to external states. The lower panel shows exactly the same dependencies but rearranged so that the internal states are associated with the intracellular states of a cell, while the sensory states become the surface states of the cell membrane overlying active states (e.g., the actin filaments of the cytoskeleton).

Under the FEP, to function as a *control system* for the organism, and ensure it remains within phenotypic states, the dynamics (i.e., *adaptive behaviour*) of the organism entails a *statistical model* (that is, a *generative* model) of *itself acting in its ecological niche*. In this framework, a generative model is a probabilistic mapping from causes in the environment (including, crucially, the actions of the organism itself) to sensory states (observations); while the generative process is the actual causal structure that generates sensed consequences from external causes that are hidden behind the MB (Friston, 2012). This inference is a necessary consequence of gradient flows that minimise variational free energy. This follows because

minimising variational free energy is – mathematically – the same as maximising the *evidence bound* in machine learning and Bayesian statistics. In this instance, the evidence is for a generative model entailed by the internal states of a creature – technically, if the internal states parameterise a posterior density over external states, then internal (and active) states will maximise the evidence for a generative model of external, niche dynamics. Active inference therefore means that the *dynamics of living systems entails a generative model*. *The optimization of this model corresponds to bounding or minimizing variational free energy,* which can be summarised as *self-evidencing* (Hohwy, 2016). This notion of self-evidencing places the MB centre stage as both an evidentiary and existential boundary.

In summary, active inference can be cast as the process that confirms and updates evidence for the statistical (generative) model that a living system entails and enacts *in existing* – thereby producing evidence for its own existence (Hohwy, 2016). This neatly covers action and perception in a folk psychology sense; because internal states cannot, in themselves, change sensory states, but they can optimise the probabilistic explanations (i.e. posterior probability distributions or Bayesian beliefs) for sensory impressions by minimising variational free energy. This has all the look and feel of perception. Conversely, active states cannot change posterior beliefs but they can change sensory states; either directly or vicariously via external states. This corresponds to action, which is informed by perception. In short, the MB will actively seek evidence for its own existence. Over time, the states of the organism (e.g., the brain) come to encode the statistical structure of causal regularities in the world, and underwrite the generative model – the *behavioural control system* (cf. Anderson, 2017; Seth, 2014) that regulates patterns of interaction with the environment. This sort of slow perception can be regarded as learning and is generally associated with plasticity of a developmental or experience-dependent sort in internal states that comprise internal structure and connectivity.

Two recent extensions of the free energy formulation will be the focus of our attention. These are variational neuroethology (VNE) (Ramstead et al., 2018) and the variational approach to niche construction (VANC) (Constant, Ramstead, et al., 2018). VNE and the VANC are about enabling the application of the FEP to phenomena within and beyond the brain. These approaches hold the promise of extending the variational (free energy) approach to the dynamics of sentient systems (i.e., systems with sensory states) across spatial and temporal scales.

3. Variational neuroethology

VNE provides a framework for modelling the dynamics of sentient systems across the spatiotemporal scales they manifest and an explanation or description of their self-organization. Formally, VNE provides a principled way of *scaling up* active inference over: (i) ensembles of MBs; and (ii) MBs of MBs. This allows us to formulate an integrative multiscale dynamics that link the partial dynamics of phenomena at each scale (see also Clark, 2017). In what follows, we speak to the multiscale aspects of MBs and then turn to ensembles of MBs that are coupled to each other.

3.1 Multiscale levels of analysis

VNE synthesises the process theory derived from the FEP (active inference) with Tinbergen's seminal four questions in biology (i.e., adaptation, phylogeny, ontogeny, and mechanism) to propose a heuristic guide to research in the sciences of life and mind. Integrating these two paradigms allows substantive insights drawn from one to inform and constrain models and research in the other – the FEP is a non-substantive principle that can be applied to any biological system in general (much like Hamilton's principle of least action), while Tinbergen's framework can provide substantive explanations for biological phenomena drawn from four complementary levels of analysis (Clark, 2004). This heuristic has already inspired an interdisciplinary theory of the human brain called the Hierarchically Mechanistic Mind (Badcock, 2012; Badcock, Friston, Ramstead, Ploeger, & Hohwy, accepted; Ramstead et al., 2018). The HMM rests on the idea that the brain is a hierarchically structured, self-organising system that has been sculpted by natural selection (Badcock, 2012; Badcock et al., 2017; Badcock, Ploeger, & Allen, 2016; Tognoli & Kelso, 2014). It suggests that the dynamics, structure, and function of the human brain instantiate an embodied, situated, *complex adaptive system* that actively minimises free energy by generating adaptive action-perception cycles via recursive interactions between hierarchically nested, functionally differentiated subsystems (Badcock, 2012; Badcock, Friston, & Ramstead, 2019). By synthesising the FEP with Tinbergen's four levels of analysis, this perspective can be reduced to four complementary research questions – What is the adaptive function of a phenotypic trait? How does it emerge from circular interactions between phylogenetic (resp. intergenerational), ontogenetic and mechanistic processes? In what ways does it instantiate the FEP? And how does it manifest in hierarchical neural dynamics? This research heuristic has already been leveraged to develop a new evidence-based theory of our capacity for depression (Badcock et al., 2017).

On the other hand, VNE entails a *multiscale ontology* for living systems as well. This ontology extends the MB formalism as a method of individuating systems under the FEP. As discussed above, a MB is a set of states that separates a system (a set of *internal* states) from *external*, non-systemic ones, which operationalises the idea that systems only exists *per se* if they are endowed with a robust form of conditional independence. According to VNE, the dynamics of living systems can be cast as active inference over *recursively nested* MBs (see Figure 2). Indeed, the MB ontology can be reiterated recursively, such that the MBs at any one scale are composed in turn of MBs at the scale above and below – which are also made of MBs, and so on, all the way up, and all the way down (Palacios et al., 2017).

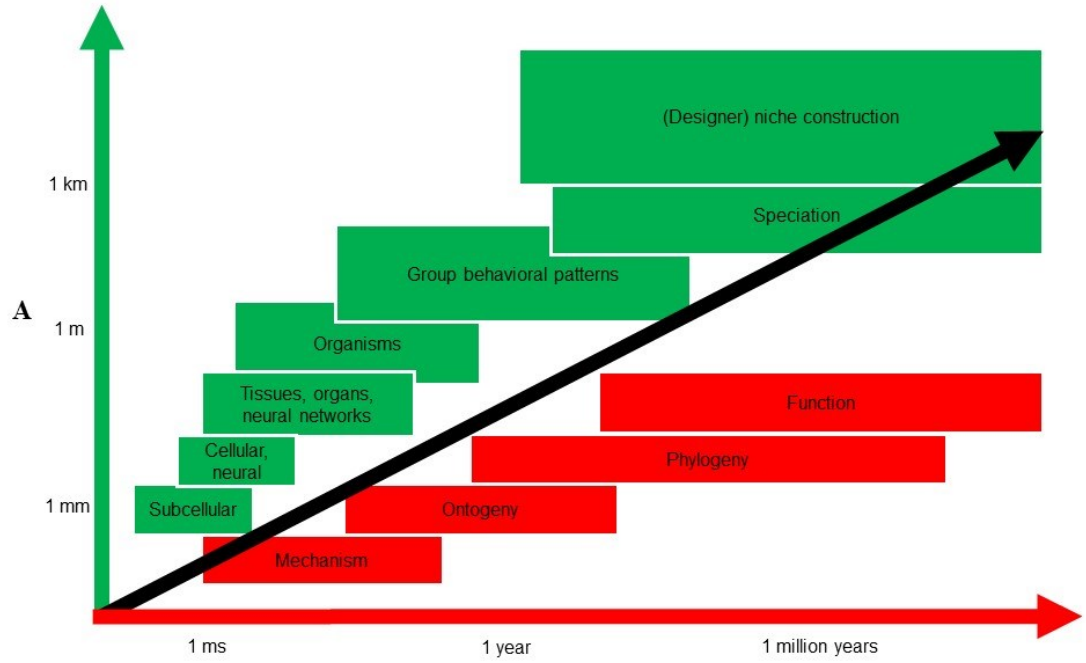


Fig. 2. *Explanatory scope of variational approach*. Variational neuroethology leverages the FEP to explain the adaptive free energy bounding dynamics of living systems across spatial and temporal scales. Here we indicate some of the scales at which these dynamics unfold. Given the intrinsic correlation between spatial and temporal scales, the phase space described here is populated mostly along its diagonal. Adapted from Sengupta, Tozzi, Cooray, Douglas, and Friston (2016).

The key notion here is that in moving from one level or scale of dynamics to the next, things not only get bigger, they also get much slower. The basic idea is that the states at one

scale constitute microscopic states that can be partitioned into an ensemble of MBs. To move to the higher scale, one treats each MB as an entity (e.g., particle) and summarises its dynamics with mixtures of blanket states that fluctuate slowly. In this multiscale setting, a (effective) state at any scale becomes the expression of an eigenmode of blanket states; namely, the principal eigenvectors of their Jacobian (i.e., rate of change of flow with respect to state). These mixtures are formally identical to *order parameters* in synergetics that reflect the amplitude of slow, unstable eigenmodes (Haken, 1983). In terms of centre manifold theory, they correspond to solutions on the slow (unstable or centre) manifold (Carr, 1981; Davis, 2006). In short, the MB of a system (or particle), at any scale, constitutes an ensemble whose order parameters subtend blanket or internal states at the scale above. Note that the constituent (microscopic) states of an ensemble are always blanket states, although their order parameters could be blanket states or internal states at the (microscopic) scale above. This follows from the fact that the only states 'that matter' are those that influence other (blanket) states. Effectively, all we are doing here is applying the (en)slaving principle, or centre manifold theorem (Haken, 1983), recursively to MBs of MBs. Figure 3 provides a schematic illustration of this recursive decomposition.

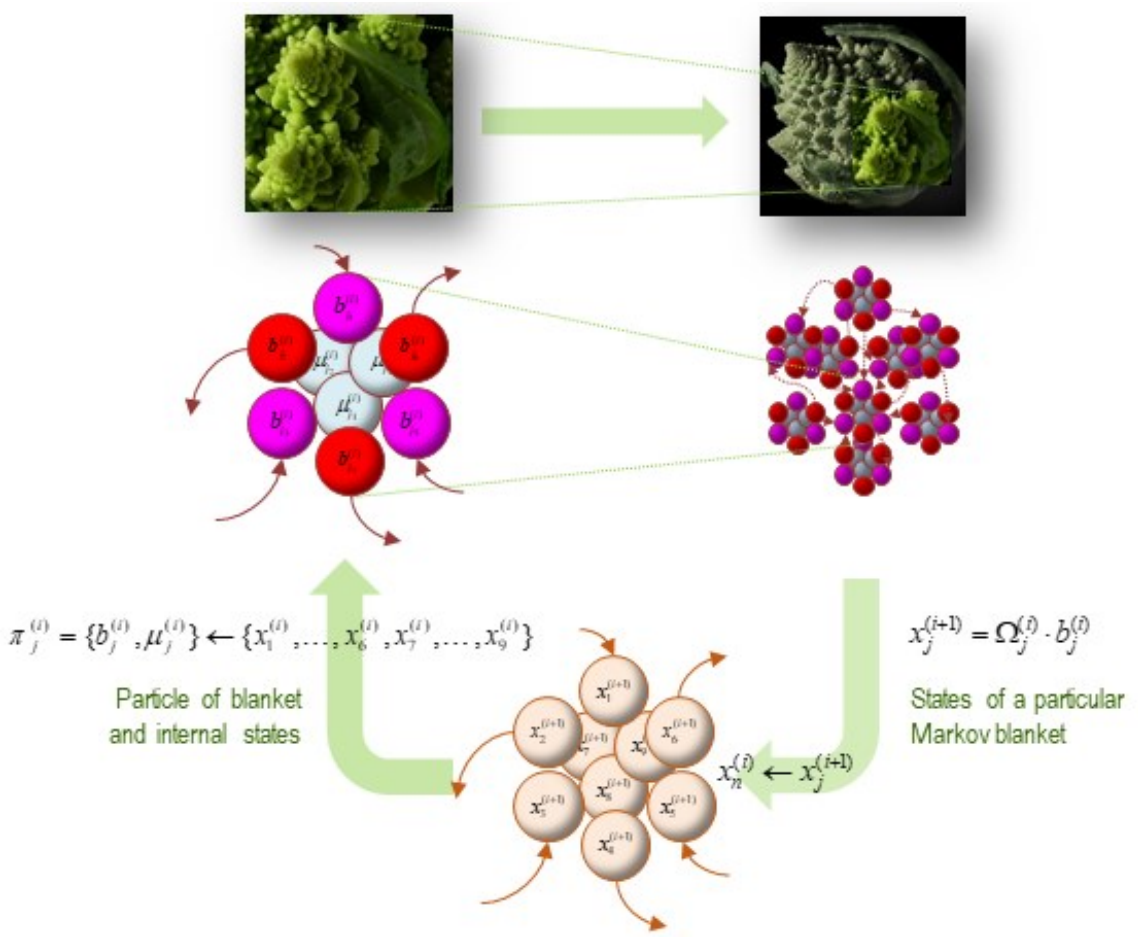


Fig. 3. *Blankets of blankets*. This schematic figure illustrates the recursive architecture by which successively larger (and slower) scale dynamics arise from subordinate levels. Starting at the bottom of the figure (lower panel) we can consider an ensemble of vector states (here nine). The conditional dependencies among these vector states then define a particular partition of into *particles* (upper panels). Crucially, this partition equips each particle with a bipartition into blanket and internal states, where blanket states comprise active (red) and sensory states (magenta). The behaviour of each particle can now be summarised in terms of (slow) eigenmodes or mixtures of its blanket states to produce vector states at the next level or scale. These constitute an ensemble of vector states and the process starts again. The upper panels illustrate the bipartition for a single particle (left panel) and an ensemble of particles; i.e., the particular partition per se (right panel). The insets on top illustrate the implicit self-similarity in moving from one scale to the next using. In this figure, $\Omega \cdot b$ denotes a linear mixture of blanket states specified by the principal eigenvectors of their Jacobian. Because the corresponding eigenvalues play the role of Lyapunov exponents, the resulting mixtures correspond to slow or unstable dynamical modes of activity.

In this multiscale framework, active inference is inherently a group activity. That is, the entire ensemble of nested MBs are bound, enslaved and constrained by dynamics at higher scales, while the lower (microscopic) scales furnish the (macroscopic) states at any given level. This construction evinces exactly the same circular causality that underlies synergetics (Frank,

2004; Haken, 1983); however, here it is generalised to a recursive hierarchy of scales – i.e., the hierarchical composition of blankets of blankets. Intuitively, the dynamics at one scale provide constraints (technically, establish probability gradients) on dynamics at other scales. Active inference destroys free energy gradients at each scale[8], under the guidance or control of a generative model at the scale above. This guidance is exerted through influences on sensory states, where circular causality means that the action of any MB in an ensemble of MBs could be involved in sensing, action or perception; depending upon its role at the superordinate scale; i.e., has a sensory, active or internal state at the level above. Please see (Palacios et al., 2017) for a worked example using simulations of biological self-assembly.

This may sound a little abstract; however, imagine you are an employee as an institution, where you transact your (microscopic) affairs with other personnel to self-evidence your prior beliefs that you are 'good at your job'. This would entail responding to corporate or institutional goals that emerge collectively (i.e., an implicit generative model at the macroscopic level). Note that your job may be homologous to an internal state at the institutional (macroscopic) level – relating only to other employees. Alternatively, you could be working on reception (i.e., a sensory state) or issuing press releases (i.e., an active state). Another example of multiscale self-organisation is provided in Figure 4 to illustrate a less anthropomorphic form of self-assembly at the cellular and molecular level. In this example, the extensive nature of variational free energy is laid bare: each system or agent that comprises the ensemble shares the same generative model. This means that the total free energy is composed in exactly the same way statisticians would accumulate statistical evidence through Bayesian belief updating with each new source of evidence[9]. The twist here is that the (sensory) evidence for each agent's model is generated by another agent. In short, multiscale ensembles that endure, in an ergodic sense, dissolve free energy gradients, thereby integrating dynamics within and between scales. This brings us to questions about coupling among MBs within a particular scale, which will be our focus for the remainder of the paper.

---

[8] The destruction of probability gradients simply means that gradient flows will seek out variational free energy minima, where the gradients disappear (Eccles & Wigfield, 2002).

[9] Bayesian belief updating treats each posterior belief – based upon some data seen so far – as the prior belief for the next set of data. This is formally equivalent to adding the log-evidence (i.e., variational free energy), because adding logs is equivalent to multiplying probabilities and evidence corresponds to the probability of the data, given a model.

3.2 Ensembles of MBs

In what follows, we look more closely at the partition of states into an ensemble of MBs – and what this means for self-organisation at any scale of dynamics. A crucial point, which enables the integration of ensemble dynamics into *adaptive behaviour*, is that variational free energy is an extensive quantity; in other words, it increases with the size or compass of the system in question. In the context of an ensemble of MBs, this means we can add the free energy of each agent or particle to describe the behaviour of the ensemble (see Figure 4).

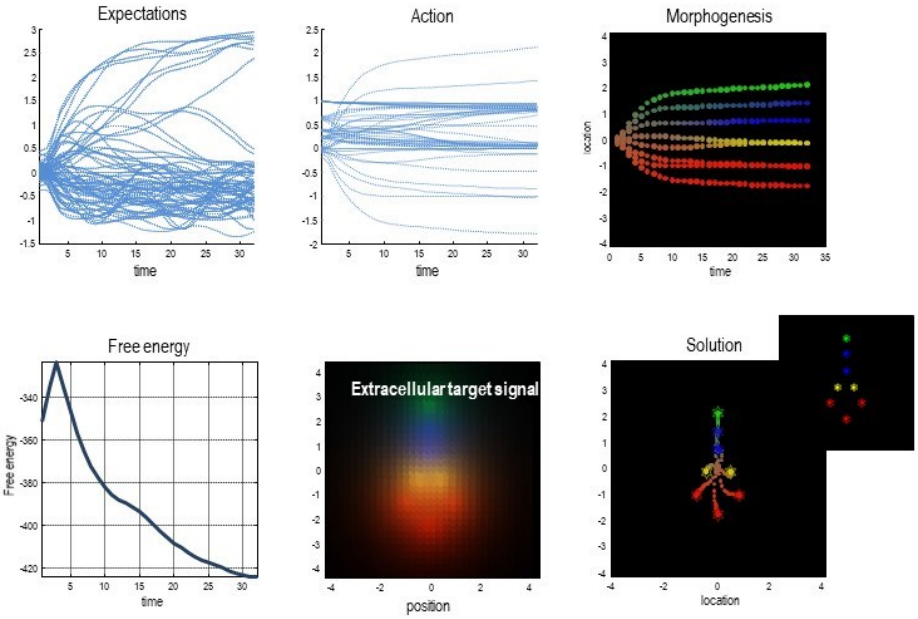


Fig. 4. *Self-assembly and active inference*. This figure shows the results of a simulation of morphogenesis under active inference reported in (Friston, Levin, Sengupta, & Pezzulo, 2015). This simulation used a gradient descent on variational free energy using a simple ensemble of eight cells; each of which had the same (pluripotential) generative model. This generative model predicted what each cell would sense and signal (chemotactically) for any given location in a 'target morphology' (lower middle panel – extracellular target signal; in other words, what each agent would expect to sense if it were at a particular location). By actively moving around, all the cells minimised their variational free energy (i.e., surprise) by inferring where they were, in relation to others. Because variational free energy is an extensive quantity, the free energy minimising arrangement of the ensemble is the target morphology. In other words, every cell has to 'find its place', at which point each cell minimises its own surprise about the signals it senses (because it knows its place) and the ensemble minimises the total free energy. The upper panels show the time courses of expectations about its place in the morphology (upper left), the

associated active states mediating migration and signal expression (upper middle) and the resulting trajectories; projected onto the first (vertical) direction – and color-coded to show differentiation (upper right). These trajectories progressively minimize total free energy (lower left panel). The lower right panel shows the ensuing configuration. Here, the trajectory is shown in small circles (for each time step). The insert corresponds to the target configuration. Please see Friston, Levin, et al. (2015) for further details.

Applying the MB formalism to sentient systems is intuitive when modelling systems whose conditional independence is maintained via literal sensorimotor loops, where the interaction between active and sensory states ranges over relatively small spatiotemporal scales. Note that every leap upward in the nested hierarchy implies a concomitant increase in spatial and temporal scale (see Figure 3), which in some cases entails an increase in the distances among the states of the system, as well as a decreased rate of change and (Bayesian) optimization. As we ascend spatiotemporal scales, it becomes increasingly unclear how the Markov boundaries are implemented. Indeed, the justification for the use of the MB formalism to describe living systems rests on the observation that if the coupling among an ensemble of dynamical subsystems is mediated by short-range forces, then the states of those remote subsystems must be conditionally independent (Friston, 2013). In summary, despite providing a framework to integrate dynamics across scales – and ensemble dynamics within a scale – VNE does not tell us how large scale ensembles self-organize so as to form higher-order MBs. In the next section, we address this issue; namely, how to draw MBs for large-scale systems like ecological niches, social ensembles, and cultural dynamics, drawing on the notion of *affordances* from the skilled intentionality framework (Bruineberg & Rietveld, 2014; Rietveld, 2008).

4. The variational approach to niche construction and the ontology of affordances

4.1. The variational approach to niche construction

Niche construction theory in evolutionary biology (e.g., Laland, Matthews, & Feldman, 2016; Lewontin, 1982; Odling-Smee, Laland, & Feldman, 2003; Stotz, 2017) argues that via their bioregulatory behaviour, living organisms (explicitly and implicitly) modify their environment, so as to steer their evolutionary trajectory, and that of other species. Arguably, then, niche construction is on a par with natural selection as a *bona fide* evolutionary force. Niche construction can be understood in two ways: in terms of selective niche construction (SNC), or

in terms of developmental niche construction (DNC) (Stotz, 2017). SNC refers to changes to the ecological niche induced by the action of organisms, and by which living systems come to modify selection pressures on themselves and other species that inhabit the niche. SNC operates on a phylogenetic time scale, and involves processes like ecological and cultural inheritance (Odling-Smee, 1988; Odling-Smee & Laland, 2011), where evolutionarily significant components of an environment are passed on from one generation to another; e.g., the remains of beaver dams are reconstructed by beavers, while the concept of a 'canoe' is passed down generations in the form of 'cognitive gadgets' (Heyes, 2018). DNC refers to the production in ontogeny of *exogenetic resources* by organisms themselves; in order to change developmental inputs, and secure the reliable and flexible reproduction of the individual life cycle (Stotz, 2017). DNC operates on the scale of development, learning, and action-perception cycles. The set of exogenetic resources that it optimizes include evolved loops of adaptive behaviour (e.g., grooming, parental care) and physical changes to the niche itself (Stotz & Griffiths, 2017).

As an example of niche construction outcome in humans, consider improvised 'desire paths'; e.g., a dirt trail in the park carved out by recurrent actions of agents in the neighbourhood. From the point of view of the FEP, such a path can function as an exogenetic resource. It encodes precise (high certainty, reliable) information about the fact that the end of the park is at the end of the trail. An agent can rely on this information to navigate the park efficiently, without having to know the layout of the neighborhood. While crossing a park might not be as essential a behaviour as grooming, from the point of view of the FEP, finding oneself in expected sensory states certainly is. Indeed, exogenetic resources of the niche such as desire paths can function as reliable indicators of surprise- and ambiguity-resolving actions, which, under the statistical conception of the phenotype entailed by the FEP, is crucial for maintaining the agent's continued existence. In short, the collective behaviour of an ensemble of agents provides a form of semiotics or a set of *possibilities for engagement with the niche* (a.k.a. *affordances*) that become relevant from the point of view of the needs and concerns of any single agent. So how can a variational ecology shed light on the role of affordance in selecting adaptive actions that emerge from ensemble dynamics?

Selecting adaptive actions requires an organism to evaluate *expected free energy* under possible action policies (Elfwing, Uchibe, & Doya, 2016). Expected free energy can be expressed in different ways; e.g., as expected energy minus entropy, or as a mixture of epistemic and pragmatic value (see Figure 5). With respect to our purposes here, expected free energy can be expressed as the *expected cost* of an outcome, given a certain action, plus the

expected *ambiguity* of the outcome (cf. Friston, Levin, et al., 2015). *Expected cost* corresponds to the *discrepancy* between outcomes conditional on a given action, and expectations or preferences about outcomes (technically, the Kullback-Leibler divergence between the two beliefs). Evaluating expected cost enables the selection of actions that bring about sensory states (observations) expected by the agent – i.e., those predicted by its generative model, and by the same token, those that once brought about, have the least deleterious or 'costly' consequences with regard to their surprisal. In turn, the ambiguity reflects an agent's expectations about the uncertainty of outcomes, dependent upon causes in the world.

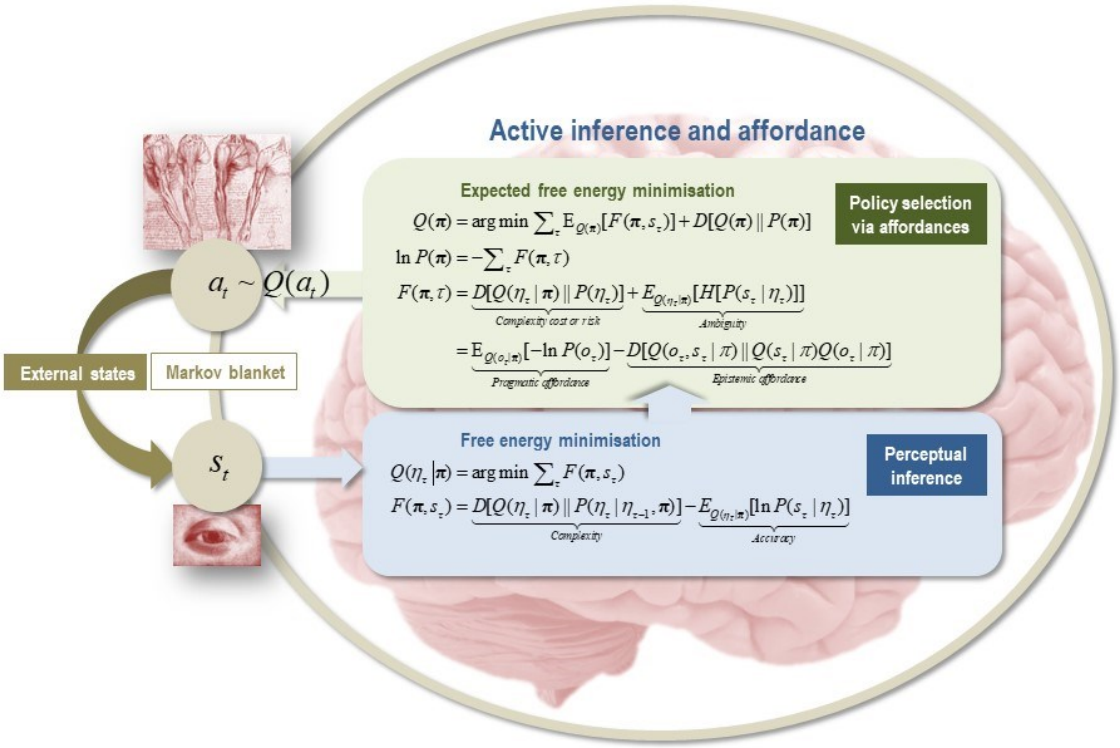


Fig. 5: *Bayesian mechanics and active inference*. This graphic summarises the belief updating scheme in the minimisation of variational free energy and expected free energy (Friston et al., 2009; Friston et al., 2016a; Kaplan & Friston, 2018). In the first step (circles on the left), discrete actions solicit a sensory outcome (i.e., in the parlance of the SIF, a *solicitation*) used to form approximate posterior beliefs about states of the world. This belief updating involves the minimisation of free energy under a set of plausible policies (blue panel – Perceptual inference). Note that free energy $F(\boldsymbol{\pi},s)$ includes Markovian dependencies among hidden states. This reflects the fact that the generative model is a Markov decision process. In the second step (green panel – Policy selection), the approximate posterior beliefs from the first step are used to evaluate *expected* free energy $F(\boldsymbol{\pi},\tau)$ and subsequent beliefs about action. These beliefs correspond to the epistemic and pragmatic *affordances* that underwrite policy

selection. Note that the free energy *per se* is a function of sensory states, given a policy. In contrast, the expected free energy is a function of the policy. The construct of *affordance* in active inference corresponds to *inferences about action* on the environment, which are selected in terms of competing policies via the minimisation of expected free energy. The variables in this figure correspond to those in Figure 1. Here, a policy $\boldsymbol{\pi}$ comprises a sequence of actions; the expression $Q(\eta \mid \boldsymbol{\pi})$ denotes beliefs about hidden states given a particular policy; and $Q(\boldsymbol{\pi})$ denotes posterior beliefs about the policy that is currently being pursued by the agent. Free energy is the difference between *complexity* and *accuracy*, while expected free energy can be decomposed into expected complexity (i.e., complexity cost or *risk*) and expected inaccuracy (i.e., *ambiguity*). Risk can be regarded as the (KL) divergence (D) between beliefs about future states under a particular policy and prior preferences about states. Ambiguity denotes the loss of a definitive mapping between external states and observed sensory states (quantified as entropy, H). Alternatively, expected free energy can be decomposed into *epistemic* and *pragmatic affordance*. Posterior beliefs about policies depend on their expected free energy. Crucially, these posterior beliefs include the free energy evaluated during perceptual inference. This has several interesting consequences from our perspective. This construction means that the agent has to infer the policy that it is currently pursuing and verify its predictions in light of sensory evidence. This is possible because the beliefs about actions that are encoded by internal states are distinct from the active states of the agent's MB. Free energy *per se* provides evidence that a particular policy is being pursued. In this scheme, agents (will appear to) entertain beliefs about their own behaviour, endowing them with what is defined as *intentionality* of goal directed behaviour under active inference. In effect, this enables agents to author their own sensorium in a fashion that has close connections with niche construction: see main text and (Bruineberg & Rietveld, 2014). See (Friston, Parr, & de Vries, 2017) for technical discussion. Figure from (Linson, Clark, Ramamoorthy, & Friston, 2018).

This is where the variational approach to niche construction (VANC) comes in. VANC exploits the expression of expected free energy as expected cost plus ambiguity, to propose that agents *upload*, as it were, much of the leg work in computing expected free energy to their self-tailored, constructed environment. More precisely, it argues that *niche construction* can be cast as the collective activity by which organisms act on their material environment to create unambiguous structure, which can be leveraged (via active inference). Lasting changes to the niche capture the fact that environmental cues function as an unambiguous indicator of the affordance of action possibilities. These can be cast in terms of *epistemic resources* that flag those actions that resolve the ambiguity associated with future observations (Friston, Rigoli, et al., 2015), while conforming to the (prior) preferences of the organism entailed by its generative model.

VANC casts DNC as the joint optimization of the niche – and its constituent ensemble of agents – over ontogeny through dense histories of active inference, and casts SNC as Bayesian model selection (Campbell, 2016; Constant, Ramstead, et al., 2018). By engaging ecologically inherited exogenetic resources of the niche, living organisms –especially those,

like humans, that depend on large-scale coordination –can reliably track, predict, and model the unfolding of causal regularities at and across large spatiotemporal scales (Clark, 2004). We will argue that this is so because niches themselves (including the organisms and their material setting) track those regularities. For instance, a complex system of agriculture that employs irrigation techniques enables groups of humans to smoothly cope with climate fluctuations that could otherwise jeopardize food production. In Section 4, we explain how the existence of higher-order, large-scale human ensembles depend on specific exogenetic resources of the niche – namely, *epistemic* resources of the kind just discussed – and how this emerges from ensembles of agents engaging in active inference.

4.2. The ontology of affordances under the variational approach

The skilled intentionality framework (SIF) (Rietveld, 2008) provides an account of the origin of *intentionality* or directed purposiveness in cognitive systems. We appeal to the SIF in offering an account of how to draw higher-order MBs; namely, by mobilizing the notion of *affordances*. The SIF recasts cognition as the engagement by organisms of the affordances that make up their local niche, thereby providing a real-time dynamics for the engagement of exogenetic (epistemic) resources that the niche affords, and enabling the stabilization of the local niche.

What are affordances, and how are they to be interpreted under the free energy formulation? The SIF defines affordances as possibilities for engagement that obtain between a set of abilities at an organism's disposal and relevant or salient features of the material environment (Bruineberg & Rietveld, 2014; Ramstead et al., 2016; Rietveld & Kiverstein, 2014; van Dijk & Rietveld, 2016). Here, 'engagement' refers to structured, skillful patterns of action and perception[10] – what we have covered under the rubric of *active inference*. In effect, the SIF provides an organism-centered dynamics that explains, concretely, what it means for an agent or group of agents and their ecological niche to engage in active inference and niche construction. Moreover, it does this by considering explicitly how the ecological niche is disclosed to the organism – as a *field of affordances*.

The SIF tells us what is special about living systems in terms of their being directed towards meaningful worlds (or strictly speaking, outcomes). Namely, it points out something special about the way living systems self-organize. The dynamics of most (non-living) self-organizing systems emerge and stabilize around an *energy gradient*, which those same

[10] These abilities are known as *policies* in active inference (please see below).

dynamics then typically resolve or consume; e.g., as a lightning bolt strikes, the charge gradient around which it organized dissipates. Unlike other self-organising systems, living organisms are unique in that they actively generate and maintain the gradients that sustain them, through adaptive actions. In other words, an organism's self-evidencing underwrites self-organisation and the very ergodicity upon which both rest (Bruineberg & Rietveld, 2014; Tschacher & Haken, 2007).

What are the gradients around which living systems organize? The variational framework suggests that these gradients are *variational free energy gradients* resolved through active inference. Consistent with the FEP, under the SIF, *affordances* are cast as *expected free energy gradients* or differences. These differences are in the expected free energy associated with the repertoire of actions or abilities available to an organism under its generative model and the learned niche (Bruineberg & Rietveld, 2014; Ramstead et al., 2016). In formulations of active inference for generative models of discrete states (i.e., Markov decision processes), these abilities correspond to the policies and their affordance is quantified in terms of the expected free energy under each policy. Given this equivalence, affordance can be decomposed into *complexity cost* and *ambiguity* (as in Figure 5). Alternatively, by rearranging its terms, expected free energy can be expressed in terms of *epistemic* (i.e., intrinsic) and *pragmatic* (i.e., extrinsic) affordance (see Friston, Rigoli, et al., 2015; Parr & Friston, 2017, for details) for details). The ensuing variational formulation of affordances uses the path integral of free energy from the current point in time to a future time point, where the only difference between expected free energy and free energy *per se* is that sensory states have yet to be realised. This means the *expectation* in *expected* free energy is over sensory states in the future, based upon posterior beliefs informed by sensory states in the past. Put simply, the best action is the next action that belongs to the policy (i.e., sequence of actions) with the greatest affordance – or the least expected free energy. This is formally related to Hamilton's principle of least Action[11] that translates here into a variational principle of greatest 'Affordance'.

For organisms to engage the affordances offered by their niche is the variational 'tissue' that connects dynamics to the niche in which those organisms exist. The expected free energy of each policy can be cast as constituting the set of affordances that an organism can entertain. Action and policy selection integrate ensemble dynamics, by minimising expected free energy directly and learning the affordances on offer from the niche by selecting courses of action (i.e.

[11] Where Action here denotes a path or time integral of free energy. The Action corresponds to the 'work done' in classical mechanics

policies) with the greatest epistemic affordance. This has close relationships with intrinsic motivation and exploration in ethology (Barto, Mirolli, & Baldassarre, 2013; Eccles & Wigfield, 2002; Oudeyer & Kaplan, 2007; Schmidhuber, 2010) where, perhaps, the most important exploration is "what can I do with my body?" (i.e., the body as niche). This is most clearly seen in development and neurorobotics (Oudeyer & Kaplan, 2007). Things get more interesting when we appreciate that the niche itself is subject to exactly the same normative principles. In other words, as each agent is trying to learn about and infer its niche, the niche – through the collective inferences, actions and material artifacts of its constituent agents – appears to learn about and predict the behaviour of each agent. This must be true, because each MB that comprises the eco-niche is itself trying to minimise variational free energy. Heuristically, this means that because every agent is trying to predict their niche, they collectively shape their field of affordances in such a way that their niche appears to infer the behaviour of its agents and therefore becomes inherently more predictable (i.e., less ambiguous).

This prompts a revision of the ontology of affordances under the SIF (Bruineberg & Rietveld, 2014). The 'field of affordances' is the set of affordances that solicit the organism at a given time. This field is constituted by expected free energy gradients that are induced by the entire ensemble. The engagement of the niche by the organism then corresponds to active inference; in the sense that these dynamics are simply the resolution of a local free energy landscape – a path of least (Hamiltonian) Action over the field of affordances. 'Solicitations' are those affordances that effectively engage the organism in action-perception loops at a given time. The 'landscape of affordances' is thus the set of affordances available in a niche at a given time. On this view, the landscape of affordances is a product of inference about "what would happen if I did that?" (Schmidhuber, 2010). Affordance is therefore an attribute of active, if counterfactual, engagement with the niche. Yet at the same time, it is a statement about the learned (and therefore lived) world.

The picture that emerges by integrating a variational approach to niche construction with the ontology of affordances could be summarised as follows. When you select actions with the greatest affordance, you learn about your niche. However, your niche comprises other MBs that must be learning about you. These can be other 'creatures like you', 'cognitive gadgets', and 'desire paths', and so on. In other words, your action on the environment constitutes sensory evidence for a niche (i.e., landscape of affordances) that is trying to model you, while the niche acts on you via sensory impressions. From the perspective of some 'Godhead' looking down on your niche, a self-organisation would emerge at a higher

spatiotemporal scale – that itself looks exactly like a self-organising, free energy minimising process. This has to be the case if your niche and all of its constituents attain some form of ergodicity, in virtue of the FEP. In other words, the network of conspecifics, their 'desire paths', and 'cognitive gadgets' would look like the internal states surrounded by a MB, separating your niche from another. We now develop this argument in the final section.

5. Variational Ecology: A physics of shared minds

How can we describe the sort of relationships that induce conditional independence among remote subsystems of ensembles (e.g., other animals as part of a larger niche)? Can we make sense of the self-organization of large-scale systems using MBs, to constitute robust and enduring sets of conditionally independent subsystems? How are MBs implemented at higher scales, and how can we define the relations among their states?

The internal states of a MB can be separated by a great deal of distance, while retaining some form of dependency. When looking at phenomena unfolding across longer temporal scales, imagining internal states as *physically isolated in a literal sense* from external states becomes conceptually constraining. Two states clearly do not need to be (and as a matter of fact cannot be) in direct contact to be part of the same system; e.g., nodes of the Internet, soldiers in a battalion. What matters is the *statistical relationship* between states; i.e., to form a MB, the right kinds of statistical relations and partitions need to be in play. In other words, ensembles sharing the same MB, whether cells or organisms, need not be 'spatially' compartmentalized, but ought to be 'statistically' or 'behaviourally' segregated; that is, some *recurrent patterns of behaviours* should govern the way different sets of internal states maintain their *conditional independence*. Practically speaking, the only thing one needs to partition states into an ensemble of MBs is their adjacency matrix. This matrix (from graph theory) encodes directed dependencies in a weighted or unweighted fashion. One can then identify a subset of internal states, their MB, and the resulting external states (Friston, 2013). This process can be repeated by identifying a second set of internal states within the external states and continuing until all external states have been exhausted. Note that, to pursue this partition into an ensemble of MBs, one only needs the adjacency matrix describing 'who is connected to whom'. This can easily be applied to differential equations describing evolutionary dynamics – or links in social media networks. See Figure 8 for an illustrative example. The question then is to determine how the *active and sensory states* are manifest, where the partition of states is dispersed over high dimensions of abstract state spaces. In other words, we need to understand what sort of 'behaviour' their states exhibit.

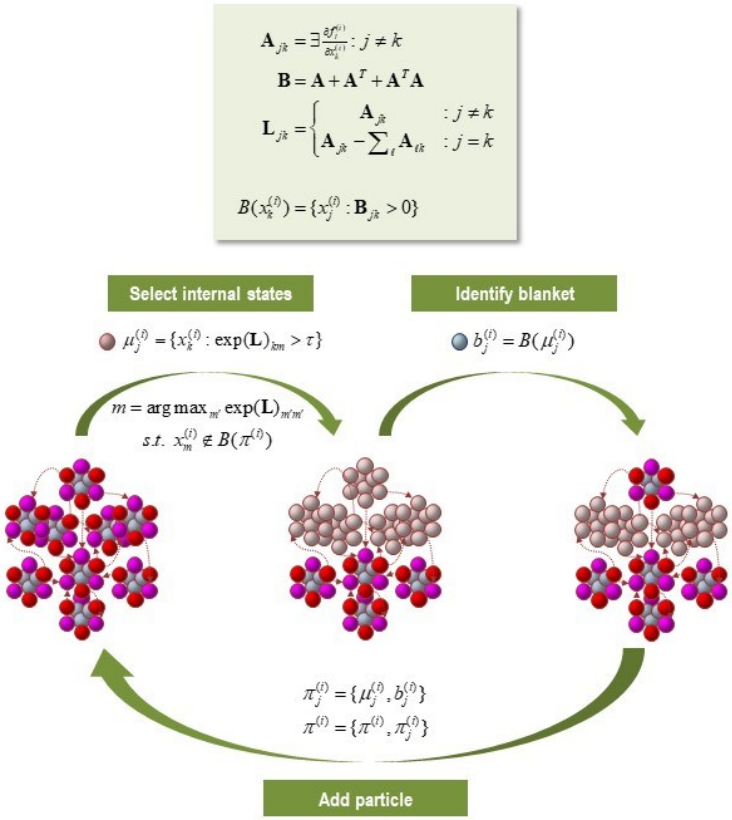


Fig. 6. *A particular partition*. This schematic figure illustrates a partition of vectors states (small coloured balls) into particles (comprising nine vectors), where each particle (π) has six blanket states (red and magenta for active and sensory states respectively) and three internal states (cyan). The upper panel summarises the operators used to create a particular partition. We start by forming an adjacency matrix that characterises the coupling between different vectors states. In this example, the adjacency matrix is based upon the Jacobian and implicitly the flow of vector states. The resulting adjacency matrix defines a MB forming matrix (**B**), which identifies the children, parents, and parents of the children. The same adjacency matrix is used to form a graph Laplacian (**L**) that is used to define neighbouring (i.e., coupled) internal states. One first identifies a set of internal states using the graph Laplacian. Here, the *j*-th subset of internal states at the level *i* are chosen based upon dense coupling with the vector state with the largest graph Laplacian. Closely coupled internal states are then selected from the columns of the graph Laplacian that exceed some threshold. In practice, the examples used later specify the number of internal states desired for each level of the hierarchical decomposition. Having identified a new set of internal states (that are not members of any particle that has been identified so far) its MB is recovered using the MB forming matrix. The internal and blanket states then constitute a new particle, which is added to the list of particles identified. This procedure is repeated until all vector states have been accounted for. In the example here, we have already identified four particles and the procedure adds a fifth (top) particle to the list of particles; thereby accounting for nine of the remaining vector states.

5.1. Markov blankets and ensemble dynamics: active and sensory states

Under the MB formalism, sensory and active states are to be interpreted in their minimal sense; that is, as *instantiating statistical relations*. What it means for a system to have active states is for it to have states the dynamics of which change as a function of some systemic quantities (namely, as a function of internal and sensory states). In turn, what it means for something to have sensory states is for it to have states that change as a function of other systemic quantities (external and active states). On this view, both sensing and acting assume a statistical definition. Indeed, one could wonder about the difference between an agent sensing things in the world with a visual saccade and a photon hitting the retina of the agent. In the parlance of the MB formalism, whether a state counts as an active or a sensory state depends on whether or not it is influenced by internal and external states in the right way (Friston, 2013). Thus, an active state is just one that is influenced by the system's defining states, and a sensory state is one that is not. Accordingly, a sensory state – at any level of description – might just mean a state that reliably covaries with, and conveys information about, some distal causes in the external system to which it is coupled (e.g., a receptionist taking calls on an external telephone line); and active states are just those states that enable the system to effect changes on the 'outside' (e.g., a public relations officer emitting press releases). An ecological niche could thus have sensory and active states (and thus, have a MB) in a sufficient and minimal statistical sense.

For a number of reasons, elements of the ontology of affordances are interesting candidates for the implementation of the sort of large scale behaviour under consideration. First, similar to the cell membrane, the sensory signals that actively engage an organism at a given time (i.e., *solicitations*) mediate the functional relationship between internal and external states of both the organism and its niche. In the parlance of the FEP, solicitations secure adaptive exchanges among the internal and external states by engaging the organism in loops of active inference that conform to internal expectations (e.g., they are perceived relative to the needs of the system), while enabling the learning of the causal structure of some external states (i.e., they are perceived relative to the actual causal structure). In this respect, similar to cell membranes, we can say that a solicitation "points both ways, to the environment and to the observer" (Gibson, 1979). In Figure 1, this 'pointing both ways' is established in virtue of the circular causality induced by the MB and ensuing active inference. This brings affordances into play, in terms of the expected free energy attributed to different policies of courses of action on the environment (see Figure 5).

Second, affordances, especially those enabled by epistemic resources, carry cultural knowledge that can be acquired in ontogeny (Costall, 1997), thereby securing the reproduction of adaptive patterned practices when transmitted across generations (Constant, Ramstead, et al., 2018). As such, affordances – especially the *cultural* affordances (Ramstead et al., 2016) that characterise a given local group – play a role in the coordination of the adaptive behaviour of members in a group over large spatial and temporal scales. Briefly, cultural affordances are the affordances with which human agents interact, and which depend on shared sets of cultural expectations, internalized through immersive practices (Ramstead et al., 2016). Affordances thus allow for adaptive behavioural self-organization among groups of agents, independent of their spatial proximity. In effect, cultural information encoded in the physical states of the ecological niche (in epistemic resources) enables the recognition and diffusion of epistemic affordances among groups over larger spatiotemporal scales (e.g., the intergenerational scale, via the passing on of affordances via cultural inheritance).

Third, defining a MB for niche systems involves the description of the conditional independence among the partitions of the system. The ontology of affordances, especially the layering it entails (e.g., organism and field, solicitation and landscape of affordances) can, in principle, be used to define the statistical compartments (partition) of the niche, and by the same token explain the conditional independence among the states of the niche.

Given these very reasons, we suggest that the *field of affordances* (and especially the solicitations that actually engage the agent at a given time) functions as the 'surface' that allows the niche to 'sense' agents via the agent's action on the niche (Bruineberg, Rietveld, Parr, van Maanen, & Friston, 2018). The agent's actions when repeated over time encode regularities about niche-agent interactions via changes to the structure of the ecological niche. In virtue of the circular causality discussed above, this structure is determined by – and indeed, determines – the affordances that underlie each agent's action. The niche then 'acts' on the agent as it is sensed by the agent. We can say that an agent will be acted upon when engaging an affordance, as the affordance is a possible action to be selected, and the selection of which will entail changes in the agent. The action of the niche thus takes the form of 'offering possibilities for engagement'. Crucially, a niche that would not offer a *variety* of possible engagements could not 'act' in any meaningful sense. The 'active' property of the niche rests on their ability to conform to the changing needs of the agent(s), that is, to solicit the agent by providing the right sort of sensory cues.

5.2. What is it that models, and what is modeled? – internal and external states

So far, we have discussed the relevant quantities to define the niche's MB. These are captured by the dynamics of the agent's field of solicitation, which involves patterns of action and sensation for the agent, which coincide with the action and perception of the niche. The key point here is that the *agent's field of affordances* (and solicitations) emerges from the dynamics of the MB of the niche, and thereby enables the niche to model sensory causal regularities, or underlying structure of the form of life it constitutes (Rietveld & Kiverstein, 2014; Wittgenstein, 1953)

Now, what are the internal states of the niche? And what are the causal regularities that they model? We suggest that internal states of the niche are a subset of the physical states of the material environment. Namely, the internal states of the niche are the physical states of the environment, which have been modified by the dense histories of different organisms interacting in their shared niche (i.e., histories of active inference). This subset of physical states comes to encode information that is used in the self-evidencing dynamics of organisms. In other words, the ecological niche becomes part of the embodied model parameters that encode the variational or recognition density that the agent uses in active inference. These model parameters encode causal regularities about agents' behaviour, which function as the external, hidden states of the niche, which it models in niche construction.

The internal states of the niche encode information about regularities of the niche-agent(s) relation (Constant, Ramstead, et al., 2018). This means that the internal states of the niche encode *organism-specific information* (e.g., affordances), not merely any changes to the physical layout. Otherwise, it would mean that the internal states would also encode random changes to the environment (e.g., volcanic eruptions and tsunamis). Hence, the ecological niche is a subset of the physical environment that the organism constructs through reiterated action over time (i.e., active inference). We can see the propensity of the environment to encode organism-specific information as the propensity to change its structure as a function of the actions of the organism (i.e., sensations of the niche) (Bruineberg et al., 2018). For instance, a grass patch has a much higher propensity to encode organisms-specific information than a sidewalk made of concrete – hence a grass patch after some time, and recurring actions, might encode a desire path (i.e., will turn into a cultural affordance). The point here is that the niche is always organisms, or species-specific, an Umwelt of sorts (Von Uexküll, 1987), whose relational structure consists of affordances, encoded through niche construction (cf. section 3).

The final quantity to define is that of the causal regularities modeled by the niche, and transcribed by its physical layout (the internal states). We have seen that for groups of

enculturated agents like humans, internal states of the niche encode affordances that pertain to group behaviour; e.g., what the desire path models is the action possibility to 'cut corners'. Now, it is a small step from this point to the one that the internal states of the niche track statistical regularities and fluctuations that underwrite the *meaning* of shared intentionality and normative group behaviour (Henrich, 2015; Tomasello & Carpenter, 2007; Tomasello, Carpenter, Call, Behne, & Moll, 2005). These regularities are highly abstract, and can only exist – *as causes – for large-scales ensembles like groups of agents*. With the example of the desire path, one can start to appreciate the continuity that obtains between the physics of self-organizing systems, the form of pragmatic engagements afforded by constructed niches, and the sort of meanings and intentions that people derive from them.

In summary, the niche transcribes regularities that pertain to *group behaviour* – they are transcribed in the physical layout of the niche – through the active and perceptual dynamics of the agent, which coincides with the perceptual and active dynamics of the niche. These same dynamics entail the landscape of action possibilities (i.e., affordances), which maintains the structural integrity of the agent-niche system – niche dynamics resolve the free energy gradients that are induced by the physical structures of organisms and their niche, and their history of dense interaction and dynamic coupling. In this sense, the robustness of patterns of shared intentionality and enaction of shared meanings is inherited from the robustness of the niche, and *vice versa*.

6. Concluding remarks

Variational ecology (VE) is a synthesis of VNE (Ramstead et al., 2018), the VANC (Constant, Ramstead, et al., 2018), and the SIF (van Dijk & Rietveld, 2016), and provides an explanation of collective purposive action and intentionality of living systems – a *physics of sentient systems*. An ecological niche 'just is' a structured set of affordances that are shared by agents, which enables its denizens to coordinate purposive action over sometimes vast spatial and temporal distances. The sort of affordances that emerge from niche construction, and that constitute large scale ensembles, carry semiotic and axiological meaning like moral states held in common (cf. Keane, 2014) (e.g., this dirt trail 'means' cutting through the park, and you shall not be late to your appointment). In effect, there is a deep sense in which affordances are what meanings are (Gibson, 1979), and a sense in which meaning is entailed by the existence of groups of agents that share a niche (cf. sense-making, De Jaegher & Di Paolo, 2007; Di Paolo & Thompson, 2014). As they engage with the affordances of their niche, ensembles of agents: (i) maintain the structural integrity of the niche through niche construction (and active

inference); and (ii) collectively enact a generative model of their relation to the niche, thereby providing an account of the physics of *intentionality*, and especially of *shared intentionality*. VE, then, is also a physics of *interacting minds*.

References

Adams, R. A., Huys, Q. J. M., & Roiser, J. P. (2016). Computational Psychiatry: towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery, and Psychiatry, 87*(1), 53-63.

Anderson, M. (2017). Of Bayes and bullets: An embodied, situated, targeting-based account of predictive processing. *Philosophy and predictive processing. Frankfurt am Main: MIND Group*.

Apps, M. A. J., & Tsakiris, M. (2014). The free-energy self: a predictive coding account of self-recognition. *Neuroscience and Biobehavioral Reviews, 41*, 85-97.

Badcock, P. B. (2012). Evolutionary systems theory: a unifying meta-theory of psychological science. *Review of General Psychology, 16*(1), 10-23.

Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017). The depressed brain: An evolutionary systems theory. *Trends in Cognitive Sciences, 21*(3), 182-194.

Badcock, P. B., Friston, K., & Ramstead, M. (2019). The Hierarchically Mechanistic Mind: A free-energy formulation of the human psyche. . *Physics of life Reviews*.

Badcock, P. B., Friston, K. J., Ramstead, M., Ploeger, A., & Hohwy, J. (accepted). The Hierarchically Mechanistic Mind: An evolutionary systems theory of the brain, mind and behavior. *Cognitive, Affective, and Behavioral Neuroscience*.

Badcock, P. B., Ploeger, A., & Allen, N. B. (2016). After phrenology: time for a paradigm shift in cognitive science. *Behavioral and Brain Sciences, 39*. doi:doi.org/10.1017/S0140525X15001557

Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or surprise? *Frontiers in Psychology, 4*(907).

Bruineberg, J., & Hesp, C. (2018). Beyond blanket terms: Challenges for the explanatory value of variational (neuro-) ethology: Comment on “Answering Schrödinger's question: A free-energy formulation” by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Bruineberg, J., Kiverstein, J., & Rietveld, E. (2016). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese*, 1-28. doi:doi:10.1007/s11229-016-1239-1

Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in human neuroscience, 8*. doi:doi.org/10.3389/fnhum.2014.00599

Bruineberg, J., Rietveld, E., Parr, T., van Maanen, L., & Friston, K. J. (2018). Free-energy minimization in joint agent-environment systems: A niche construction perspective. *Journal of Theoretical Biology, 455*, 161-178.

Campbell, J. O. (2016). Universal Darwinism as a process of Bayesian inference. *Frontiers in Systems Neuroscience, 10*.

Carr, J. (1981). *Applications of Centre Manifold Theory*. Berlin: Springer-Verlag.

Clark, A. (2004). *Natural-born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford: Oxford University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181-204.

Clark, A. (2015). *Surfing uncertainty: prediction, action, and the embodied mind*. Oxford: Oxford University Press.

Clark, A. (2017). How to knit your own Markov blanket. In T. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing*. Frankfurt am Main: MIND Group.

Constant, A., Bervoets, J., Hens, K., & Van de Cruys, S. (2018). Precise worlds for certain minds: An ecological perspective on the relational self in autism. *Topoi*.

Constant, A., Ramstead, M., Veissière, S., Campbell, J., & Friston, K. (2018). A variational approach to niche construction. *Journal of the Royal Society Interface*.

Corlett, P. R., & Fletcher, P. C. (2014). Computational psychiatry: a Rosetta Stone linking the brain to mental illness. *The Lancet Psychiatry, 1*(5), 399-402.

Costall, A. (1997). The meaning of things. *Social Analysis: The International Journal of Social and Cultural Practice, 41*(1), 76-85.

Davis, M. J. (2006). Low-dimensional manifolds in reaction- diffusion equations. 1. Fundamental aspects. *The Journal of Physical Chemistry A, 110*(16), 5235-5256.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The helmholtz machine. *Neural computation, 7*(5), 889-904.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences, 6*(4), 485-507.

Di Paolo, E. A., & Thompson, E. (2014). The enactive approach. In L. E. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 68-78). London: Routledge.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology, 53*, 109-132.

Elfwing, S., Uchibe, E., & Doya, K. (2016). From free energy to expected energy: Improving energy-based value function approximation in reinforcement learning. *Neural Networks: The Official Journal of the International Neural Network Society, 84*(17-27).

Evans, D. J., & Searles, D. J. (2002). The Fluctuation Theorem. *Advances in physics, 51*(7), 1529-1585.

Fabry, R. E. (2017). Betwixt and between: the enculturated predictive processing approach to cognition. *Synthese*, 1-36.

Frank, T. D. (2004). *Nonlinear Fokker-Planck Equations: Fundamentals and Applications*. Berlin: Springer.

Friston, K. (2011). Embodied inference: or ‘‘I think therefore I am, if I am what I think’’. In W. Tschacher & C. Bergomi (Eds.), *The implications of embodiment: Cognition and communication* (pp. 89-125). Exeter, UK: Imprint Academic.

Friston, K. (2012). A free energy principle for biological systems. *Entropy, 14*(11), 2100-2121.

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

Friston, K., & Ao, P. (2011). Free energy, value, and attractors. *Computational and mathematical methods in medicine, 2012*.

Friston, K., Daunizeau, J., & Kiebel, S. (2009). Reinforcement learning or active inference? *PloS one, 4*(7), e6421.

Friston, K., Daunizeau, J., Kilner, J., & Kiebel, S. (2010). Action and behavior: a free-energy formulation. *Biological cybernetics, 102*(3), 227-260.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2016a). Active inference and learning. *Neuroscience & Biobehavioral Reviews, 68*, 862-879.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive neuroscience, 6*(4), 187-214.

Friston, K. J. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences, 360*(1456), 815-836.

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. (2016b). Active inference: a process theory. *Neural computation, 29*, 1-49.

Friston, K. J., & Frith, C. D. (2015). Active inference, communication and hermeneutics. *cortex, 68*, 129-143.

Friston, K. J., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris, 100*(1), 70-87.

Friston, K. J., Levin, M., Sengupta, B., & Pezzulo, G. (2015). Knowing one's place: a free-energy approach to pattern regulation. *Journal of the Royal Society Interface, 12*(105).

Friston, K. J., Parr, T., & de Vries, B. (2017). The graphical brain: belief propagation and active inference. *Network Neuroscience, 1*(4), 381-414.

Friston, K. J., Rosch, R., Parr, T., Price, C., & Bowman, H. (2018). Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews, 90*, 486-501.

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese, 159*(3), 417-458.

Gibson, J. J. (1979). *The ecological approach to visual perception*: Psychology Press.

Haken, H. (1983). *Synergetics: An introduction. Non-equilibrium phase transition and self-selforganisation in physics, chemistry and biology*. Berlin: Springer-Verlag.

Henrich, J. (2015). *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton, NJ: Princeton University Press.

Heyes, C. (2018). *Cognitive gadgets: the cultural evolution of thinking*: Harvard University Press.

Hohwy, J. (2014). *The predictive mind*. Oxford: Oxford University Press.

Hohwy, J. (2016). The self-evidencing brain. *Noûs, 50*(2), 259-285.

Huang, G. T. (2008). Essence of thought. *New Scientist, 198*(2658), 30-33.

Kaplan, R., & Friston, K. J. (2018). Planning and navigation as active inference. *Biological cybernetics*, 1-21.

Keane, W. (2014). Affordances and reflexivity in ethical life: An ethnographic stance. *Anthropological Theory, 14*(1), 3-26.

Kiebel, S. J., & Friston, K. J. (2011). Free energy and dendritic self-organization. *Frontiers in Systems Neuroscience, 5*.

Kirmayer, L., & Ramstead, M. (2017). Embodiment and Enactment in Cultural Psychiatry. In *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*: MIT Press

Kirmayer, L. J. (2018). Ontologies of life: From thermodynamics to teleonomics. Comment on “Answering Schrödinger’s question: A free-energy formulation” by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in neurosciences, 27*(12), 712-719.

Laland, K., Matthews, B., & Feldman, M. W. (2016). An introduction to niche construction theory. *Evolutionary Ecology, 30*, 191-202.

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in human neuroscience, 302*.

Lewontin, R. C. (1982). Organism and environment. In H. C. Plotkin (Ed.), *Learning, development, and culture: Essays in evolutionary epistemology*. New York: Wiley.

Linson, A., Clark, A., Ramamoorthy, S., & Friston, K. (2018). The active inference approach to ecological perception: general information dynamics for natural and artificial embodied cognition. *Frontiers in Robotics and AI, 5*, 21.

Metzinger, T., & Wiese, W. (Eds.). (2017). *Philosophy and predictive processing*. Frankfurt am Main: MIND Group.

Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). *Niche construction: The neglected process in evolution*. Princeton: Princeton University Press.

Odling-Smee, J. (1988). Niche constructing phenotypes. In H. Plotkin (Ed.), *The role of behavior in evolution*. Cambridge, MA: MIT Press.

Odling-Smee, J., & Laland, K. N. (2011). Ecological inheritance and cultural inheritance: what are they and how do they differ? *Biological Theory, 6*(3), 220-230.

Oudeyer, P.-Y., & Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in neurorobotics, 1*(6).

Palacios, E., Razi, A., Parr, T., Kirchhoff, M., & Friston, K. (2017). Biological self-organisation and Markov blankets. *bioRxiv*, 227181.

Parr, T., & Friston, K. J. (2017). Working memory, attention, and salience in active inference. *Scientific reports, 7*(1), 14678.

Ramstead, M., Badcock, P. B., & Friston, K. (2018). Answering Schrödinger's question: A free-energy formulation. *Physics of life Reviews, 24*, 1-16.

Ramstead, M., Veissière, S., & Kirmayer, L. (2016). Cultural affordances: scaffolding local worlds through shared intentionality and regimes of attention. *Frontiers in Psychology, 7*.

Rao, R. P., & 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.

Redish, A. D., & Gordon, J. A. (Eds.). (2016). *Computational psychiatry: New perspectives on mental illness*. Cambridge, MA: MIT Press

Rietveld, E. (2008). Special section: The skillful body as a concernful system of possible actions phenomena and neurodynamics. *Theory & Psychology, 18*, 341-363.

Rietveld, E., & Kiverstein, J. (2014). A rich landscape of affordances. *Ecological Psychology, 26*(4), 325-352.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development, 2*(3), 230-247.

Seifert, U. (2012). Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics, 75*(12).

Sengupta, B., Tozzi, A., Cooray, G. K., Douglas, P. K., & Friston, K. J. (2016). Towards a neuronal gauge theory. *PLOS Biology, 14*(3).

Seth, A. K. (2014). The cybernetic brain: from interoceptive inference to sensorimotor contingencies. In T. Metzinger & J. M. Windt (Eds.), *Open MIND* (pp. 1-24). Frankfurt am Main: MIND Group.

Stearns, S. C. (1989). The evolutionary significance of phenotypic plasticity phenotypic sources of variation among organisms can be described by developmental switches and reaction norms. *Bioscience,, 39*(7), 436-445.

Stotz, K. (2017). Why developmental niche construction is not selective niche construction: and why it matters. *Interface focus, 7*(5), 20160157.

Stotz, K., & Griffiths, P. E. (2017). A developmental systems account of human nature. In T. Lewens & E. Hannon (Eds.), *Why We Disagree About Human Nature*. Oxford & New York: Oxford University Press.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie, 20*, 410-433.

Tognoli, E., & Kelso, J. A. S. (2014). The metastable brain. *Neuron, 81*(1), 35-48.

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science, 10*(1), 121-125.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*(5), 675-691.

Tschacher, W., & Haken, H. (2007). Intentionality in non-equilibrium systems? The functional aspects of self-organized pattern formation. *New Ideas in Psychology, 25*(1), 1-15.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychological review, 121*(4), 649-675.

van Dijk, L., & Rietveld, E. (2016). Foregrounding sociomaterial practice in our understanding of affordances: The Skilled Intentionality Framework. *Frontiers in Psychology, 7*(1969).

Von Uexküll, T. (1987). The sign theory of Jakob von Uexküll. In *Classics of semiotics* (pp. 147-179): Springer.

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford: Oxford University Press.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.

8.4. Chapter 4: Multiscale Integration: Beyond Internalism and Externalism

Original publication details:

Ramstead, M. J. D., Kirchhoff, M. D., Constant, A., & Friston, K. J. (2019). Multiscale integration: Beyond internalism and externalism. *Synthese*. doi.org/10.1007/s11229-019-02115-x

Authors:

Maxwell J. D. Ramstead[1,2,3*]
Michael D. Kirchhoff[4]
Axel Constant[3,5]
Karl J. Friston[3]

Affiliations:

1. Department of Philosophy, McGill University, Montreal, Quebec, Canada.
2. Division of Social and Transcultural Psychiatry, Department of Psychiatry, McGill University, Montreal, Quebec, Canada.
3. Wellcome Trust Centre for Neuroimaging, University College London, London, UK, WC1N3BG.
4. Department of Philosophy, Faculty of Law, Humanities and the Arts, University of Wollongong, Wollongong, Australia
5. Amsterdam Brain and Cognition centre, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands.
* Corresponding author

Abstract:

We present a multiscale integrationist interpretation of the boundaries of cognitive systems, using the Markov blanket formalism of the variational free energy principle (FEP). This interpretation is intended as a corrective for the philosophical debate over internalist and externalist interpretations of cognitive boundaries; we stake out a compromise position. We first survey key principles of new radical (extended, enactive, embodied) views of cognition.

We then describe an internalist interpretation premised on the Markov blanket formalism. Having reviewed these accounts, we develop our positive multiscale account. We argue that the statistical seclusion of internal from external states of the system – entailed by the existence of a Markov boundary – can coexist happily with the multiscale integration of the system through its dynamics. Our approach does not privilege any given boundary (whether it be that of the brain, body, or world), nor does it argue that all boundaries are equally prescient. We argue that the relevant boundaries of cognition depend on the level being characterised and the explanatory interests that guide investigation. We approach the issue of how and where to draw the boundaries of cognitive systems through a multiscale ontology of cognitive systems, which offers a multidisciplinary research heuristic for cognitive science.

Keywords:
Boundaries of cognition; Variational free energy principle; Externalism; Internalism; Enactive cognition; Embodied cognition; Markov blankets

Acknowledgements:
This research was undertaken thanks in part to funding from the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative (M. J. D. Ramstead), the Social Sciences and Humanities Research Council of Canada (M. J. D. Ramstead), the Australian Research Council (M. D. Kirchhoff – DP170102987), and by a Wellcome Trust Principal Research Fellowship (K. J. Friston – 088130/Z/09/Z). We thank Jelle Bruineberg, Laurence Kirmayer, Jonathan St-Onge, Samuel Veissière, and two anonymous reviewers for helpful discussions and comments.

In the previous chapter,
I developed a framework for cognitive ecology, called variational ecology (VE), which is formulated to model the manner in which hierarchically structured cognitive systems – nested living systems from cells to societies – construct, and align themselves with, their ecological niche. Taken together with the results of the two previous chapters, VE provides the cognitive sciences with a fully generalisable metatheory of adaptive cognitive dynamics in living systems. In this chapter, I draw the main philosophical consequences of adopting this framework for the philosophical issue of where to draw the boundaries of cognitive systems. I argue that VNE and VE entail that the boundaries of living systems are multiple, nested, and relative to our explanatory interests. The combined framework afforded by VNE and VE

shows that to explain the adaptive behaviour of any cognitive system, we must consider the many nested boundaries of the systems that constitute the system of interest. The chapter also offers an argument for methodological pluralism based on VE and VNE.

> What we think cognition is depends on what our theoretical commitments suggest can be explained. The question facing the field is not "Which approach is true?" but "Which approach gives us the best scientific leverage?"
> (Hutchins, 2010, pp. 706-707)

1. Introduction

Over two decades ago, in 1991, Francisco Varela and colleagues articulated a general idea that now underlies what might be called *radical views on cognition*; namely, enactive, embodied, and extended approaches to cognition. According to proponents of the *enactive approach*, "cognition is … the enactment of a world and a mind on the basis of a history of the variety of actions that a being in the world performs" (Varela, Thompson, & Rosch, 1991, p. 9). Since Varela and colleagues, philosophers and scientists have addressed the role of *embodied activity* in cognition and the degree to which our cognitive capacities are realised partly by elements of our embedding environment. Philosophers especially have been considering what embodied, enactive, and extended accounts have to teach us about the *boundaries of cognitive systems*.

Here, we focus on making explicit a description of the boundaries of cognitive systems that we think follows from taking seriously the enactive, embodied, and extended nature of cognition. This is the idea that *the boundaries of cognitive systems are nested and multiple – and that, with respect to its study, cognition has no fixed or essential boundaries* (Clark, 2017; Kirchhoff, 2018c; Kirchhoff & Kiverstein, 2019; Kirchhoff, 2012; Stotz, 2010; Sutton, 2010).

This idea is far from the accepted view in the philosophy of mind and cognition. Indeed, it is common for researchers from different fields of study – e.g., neuroscience and the philosophy of neuroscience (Hohwy, 2014; Seth, 2014), embodied cognition (Gallagher, 2006; Noë, 2004), ecological psychology (Gibson, 1979), and anthropology (Ingold, 2001) – to infer that there is a *uniquely defining boundary* or unit of analysis from which best to understand and investigate cognition. In its more extreme forms, one might call this position *essentialism* about the boundaries of cognition. Views stressing that cognition has a unique and privileged boundary take many forms. Some argue that cognitive activity is essentially realised by states of the brain. Others argue that cognition is best conceived of as forms of embodied activity. Others still prefer to study cognition "in the wild," in terms of the patterning of cultural practices and construction of cognitive niches.

The claim that the boundaries of cognition are nested and varied runs counter to any of these brain-based, embodied, and/or ecological, environmental assumptions about the boundaries of cognition, for it does not privilege the brain, the body, or the environment. Nor do we consider the brain, body and environment as equally important, as some in the enactivist tradition have proposed (Hutto & Myin, 2013). This is the Equal Partner Principle of radical enactivism. It states that the contributions of the brain to cognition should not be prioritised over those of the body and the environment. Even if there is something correct about this claim – that one should not a priori privilege the brain in explanations of cognition – there is also something problematic about this principle; namely, that on some occasions it will turn out to be incorrect, as privileging the brain will be required to explain some phenomena under consideration.

Where to draw the *scientifically relevant boundaries* will depend both on the nature of the phenomenon being investigated and on our explanatory interests (Clark, 2017; Hutchins, 1995). By standing on the shoulders of theorists that take seriously the idea that cognitive boundaries are not singular but nested and varied, we reject all views assuming there to be unique and privileged boundaries for cognitive systems, and stake out a compromise position between (in our view) the overly coarse-grained distinction between internalism and externalism about the boundaries of cognition.

Our argument takes the form of a *multiscale integrationist* formulation of the boundaries of cognition based on the variational free energy principle (henceforth FEP). This principle casts cognition and action in terms of quantities that change to minimise free energy expected under action policies. As we discuss in the second section of this paper, we use the FEP because free energy and its expectation can be broadly construed as metrics of cognitive activity that transcend specific spatial and temporal scales (Friston, Levin, Sengupta, & Pezzulo, 2015; Kirchhoff, 2015; Kirchhoff, Parr, Palacios, Friston, & Kiverstein, 2018; Ramstead, Badcock, & Friston, 2018a; Ramstead, Constant, Badcock, & Friston, 2019). This allows us to cast the boundaries of cognition as assembled and maintained in an informational dynamics across multiple spatial and temporal scales. Crucially, we shall show that this multiscale application of the FEP implies both ontological and methodological pluralism.

We cast *ontological pluralism* in terms of a *multiscale formal ontology of cognitive systems*. In the sense we are using the term, to produce a *formal ontology* means to use a mathematical formalism to answer the questions traditionally posed by metaphysics; i.e., what does it mean to be a thing that exists, what is existence, etc. Our formal ontology is

effectively in the same game as statistical physics, in that it treats as a system sets of states that evince a robust enough form of conditional independence.

This ontology implies, that any given cognitive system has a plurality of boundaries relevant to their scientific study; namely, the boundaries of its relevant subsystems. Our claim is that which among these are the *most relevant* will depend on the phenomenon being studied and the explanatory interests of researchers. Some of these boundaries are internal to the systems – these are boundaries of relevant *subsystems* nested in the whole system or organism (e.g., cells, ensemble of cells, organs, etc.); other boundaries separate the organism from its external environment (e.g., the skin membrane); and others still extend outwards to include the organism and external, worldly states (e.g., constructed niches and patterned cultural practices).

The claims we are making about the boundaries of cognitive systems are ontological. We are using a mathematical formalism to answer questions that are traditionally those of the discipline of ontology, but crucially, we are not deciding any of the ontological questions in an a priori manner. The Markov blankets are a result of the system's dynamics. In a sense, we are letting the biological systems carve out their own boundaries in applying this formalism. Hence, we are endorsing a dynamic and self-organising ontology of systemic boundaries.

Furthermore, this ontological pluralism implies *methodological pluralism* under the FEP. The FEP can be used as a *methodological heuristic* for interdisciplinary research, which in turn allows scientists to privilege various boundaries of a nested cognitive system, depending on their specific explanatory interests. The FEP is *not* a theory of everything; it does not, on its own, provide an explanation of the systemic processes that constitute living systems (Ramstead, Badcock, & Friston, 2018b). Rather, it is a principle that coordinates and constrains the kind of explanations deployed when one is addressing how expected free energy minimisation occurs across many different spatial and temporal scales; which call for complementary explanations in terms of, e.g., neuroscience (Friston, 2010), embodied cognition (Allen & Friston, 2016), ecological psychology (Bruineberg & Rietveld, 2014; Ramstead et al., 2019), and niche construction (Constant, Ramstead, Veissière, Campbell, & Friston, 2018; Hesp et al., 2019).[12]

We approach this multiscale, integrationist view of the boundaries of cognition by focusing on the *Markov blanket formalism*, which underwrites the FEP (see Figure 1 for a detailed technical explanation). This formalism allows us to individuate a system by

[12] We do not intend this list of relevant disciplines to be exhaustive.

demarcating its boundaries in a statistical sense. Intuitively, for a thing to exist, it must evince some form of conditional independence from the system in which it is embedded. Markov blankets operationalise this intuition. In more technical terms, a Markov blanket induces a statistical partitioning between internal (systemic) and external (environmental) states, where environmental states can be associated with neuronal, bodily, or worldly states depending on the relevant partitioning of the system in question. The Markov blanket itself comprises a bipartition into active and sensory states, which mediate exchanges between systemic and environmental (neuronal, bodily, worldly) states. Importantly, the presence of a Markov blanket shields or insulates internal states from the direct influence of external states. This follows from the partitioning rule of Markov blankets, according to which internal states can influence external states via active states, and external states can influence internal states via sensory states. Hence, the Markov blanket formalism shows that internal and external states are 'hidden' (i.e., conditionally independent) from one another in virtue of the existence of a Markov blanket, thus providing the statistical means by which to delineate the boundaries of a biological and/or cognitive system.

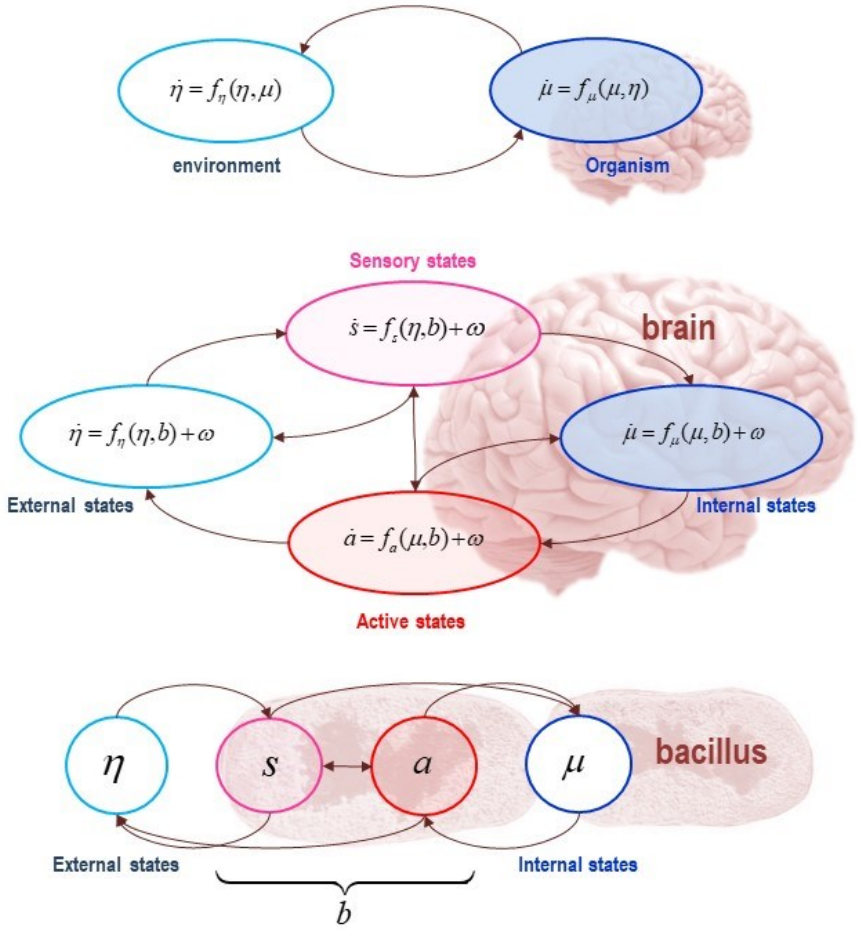


Fig 1. *The Markov blanket and active inference*. A Markov blanket is a set of states that enshrouds or statistically isolates internal states from external or hidden states. This figure depicts the partition of the states into internal (μ) and external states (η). In the parlance of graph theory, the Markov blanket is a set of nodes that shields the internal states (or nodes) from the influence of external states; in the sense that internal states can only be affected by external states indirectly, via the blanket states (Friston, Parr, & de Vries, 2017). Internal and external states are therefore separated, in a statistical sense, by the Markov blanket (b), which itself comprises sensory (s) and active states (a) – defined as blanket states that are and are not influenced by external states respectively. The top panel schematises the relations of reciprocal causation that couple the organism to its ecological niche, and back again. Internal states of the organism change as a function of its current state (μ) and the state of its niche (η), which is expressed in terms of a flow $f_\mu(\mu,\eta)$ with random fluctuations. Reciprocally, states of the niche change over time as a function of the current state of the environment and the organism, again, specified in terms of a flow $f_\eta(\eta,\mu)$ with random fluctuations. The self-organisation of internal states in this scheme corresponds to perception. Active states couple internal states back to states of the niche, and so correspond to the actions of an organism. Given the anti-symmetric conditional dependencies entailed by the presence of the Markov blanket, the dynamics of the niche, too, can be expressed as a gradient flow of a free energy functional of external and blanket states. The lower panel depicts the dependencies as they would apply to a unicellular organism. In this panel, the internal states are associated with

the intracellular states of a cell, the sensory states are associated with surface states of the cell membrane, and the active states are associated with the actin filaments of the cytoskeleton. Adapted from Constant, Ramstead, et al. (2018).

We accept that the Markov blanket formalism can be used to delineate the boundaries of cognitive systems (cf. Hohwy, 2016; Kirchhoff & Kiverstein, 2019). We shall argue that cognition involves dynamics (i.e., the Bayesian mechanics of active inference) that ensure adaptiveness, which straddle across and integrate such boundaries. We call this position *multiscale integration*. We argue that the FEP can accommodate a multiscale integrationist account of the boundaries of cognitive systems. We therefore argue that the *inferential seclusion* of internal states and external states, given by the Markov blanket formalism, can coexist with *existential integration* through active inference; justifying the view that the boundaries of cognition are *nested* and *multiple*.

The structure of this paper is as follows. In the next (second) section, we review the FEP and active inference. In the third section, we survey key principles of new radical – extended, enactive, embodied – views of cognition, with a focus on enactive views in particular. We then describe a brain-based argument for the boundary of cognitive systems premised on the Markov blanket formalism – and the FEP – that pushes back again these radical views of cognition. In the fourth section, we develop our positive proposal for a multiscale account of the FEP. We argue that the encapsulation or statistical seclusion entailed by the Markov boundary is reiterated at every hierarchical description of living systems; from the single cell, to organs, to individuals, and all the way out to coupled organism-environment systems – all of which can be cast as having their own Markov blanket. We also argue that the organism and niche are coupled to one another through active inference.

In this sense, our argument owes much to Clark (2017). Clark sets out the idea of organisms having temporally extended Markov blankets, the boundaries of which reach all the way down to DNA and all the way up to individual organisms and their respective niches. Our focus, however, is different from Clark's, in two ways. First, we make explicit that this view of the Markov blanketed cognitive system implies two forms of pluralism, ontological and methodological; and second, we emphasize that active inference entails adaptive phenotypes, cultural practices, and niche construction; the joint phenotype of the organism (including states of its adapted niche) encodes information that, at least in some cases, is as

important as that encoded by states of the brain to explain adaptive behaviour. We conclude by considering future research directions for approaching systemic organisation through a multiscale ontology of cognitive systems and a multidisciplinary research heuristic for cognitive science.

2. A variational principle for living systems

2.1. The variational free energy formulation

Organisms find themselves, more often than not, in a bounded set of characteristic states. We can cast this set of states, in which the organism is *most likely* to find itself, as its overall *phenotypical states and traits*; namely, the repertoire of measurable functional and physiological states, as well as morphological traits, behavioural patterns, and the adapted ecological niches that characterizes it as 'the kind of organism that it is' (Kirchhoff et al., 2018; Ramstead et al., 2018a). From this statistical perspective, the question of how organisms remain alive can be recast as the question of how they maintain themselves in phenotypic states.

Remarkably, organisms resist entropic erosion by simply limiting the dispersion of states that they occupy during their lifetime. The variational free energy principle (FEP) provides a formal description of this anti-entropic feat. The FEP casts the functioning of biological systems of any kind, including their different psychological profiles, in terms of a single imperative: to minimise *surprise* (aka surprisal or self-information). The concept of surprise does not refer to the psychological phenomenon of being surprised. It is an information-theoretic notion that measures how uncharacteristic or unexpected a particular sensory state is, where *sensory* states can be caused by *external* worldly (and bodily) states.

A key premise of the FEP is that cognitive systems cannot estimate surprise directly and therefore must work to reduce an *upper bound* on surprise, which they can track; namely, variational free energy. In other words, surprise cannot be evaluated directly because this would entail to name all possible ways in which some sensations could have been caused. However, variational free energy can be evaluated given a generative model of how sensations were caused. Because variational free energy is (by construction) always greater than surprise, minimising free energy implicitly minimises surprise (see Figure 2). One can think of variational free energy as a *guess* or *approximation to surprise*, whose accuracy can be finessed through perception; namely, the dynamics of a system's internal states. This

perceptually crafted approximation to surprise can now be minimised by action; namely, the dynamics of a system's active states.

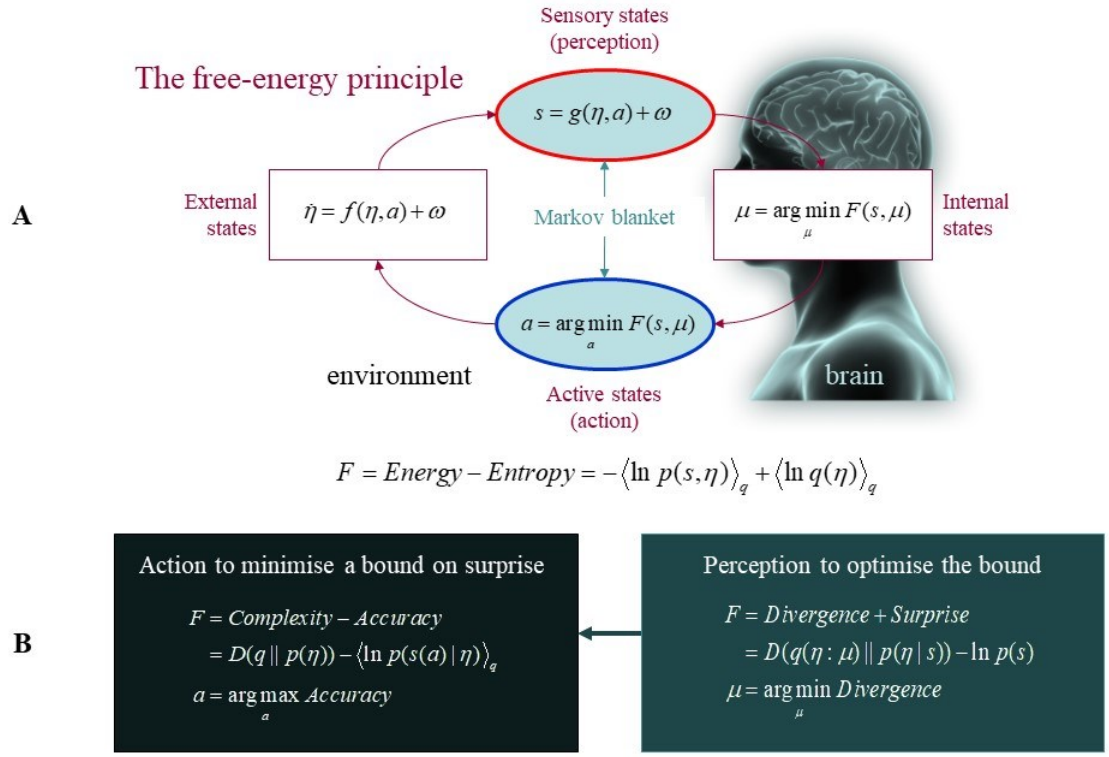


Fig. 2. The free energy principle and self-evidencing. Upper panel: Depiction of the quantities that define an agent engaging in active inference and its coupling to its ecological niche or environment. These are the internal states of the agent (μ), sensory input $s = g(\eta, a) + \omega$, and action a. Action and sensory input describe exchanges between the agent and its world; in particular, action changes how the organism samples its environment. The environment is described by equations of motion, $\dot{\eta}$ $) + \omega$, that specify the (stochastic) dynamics of (hidden) states of the world η. Here, ω denote random fluctuations. The free energy (F) is a function of sensory input and a probabilistic belief $q(\eta : \mu)$ that is encoded by internal states. Changes in active states and internal states both minimise free-energy and, implicitly, self-information. Lower panel: Depiction of alternative expressions for the variational free-energy, which clarify what its minimisation entails. With regards to action, free-energy can only be minimised by increasing the *accuracy* of sensory data (i.e., the selective sampling of predicted data). Conversely, the optimisation of internal states through perception makes the probability distribution encoded by internal states an approximate conditional density on the causes of sensory input (by minimising a Kullback-Leibler divergence D between the approximate and true posterior density). This optimisation tightens the free-energy bound on self-information and enables the creature to avoid surprising sensations through adaptive action (because the divergence can never be less than zero). With regards to the selection of actions that minimise the *expected free energy*, the expected divergence becomes (negative) *epistemic value* or *salience*, and the expected surprise becomes (negative) *extrinsic value*; which is the expected

likelihood that prior preferences are indeed realised as a result of the selected action. See Friston, FitzGerald, Rigoli, Schwartenbeck, and Pezzulo (2017) for a full description of the free energy expected following an action. Adapted from Ramstead et al. (2018a).

In a nutshell, the FEP tells us that cognitive systems can estimate and thereby avoid surprise, on average and over time, by working to suppress a variational bound on surprise. Crucially, this free energy bound is exactly the same quantity used in Bayesian statistics to optimise (generative) models of data. In this setting, negative free energy is known as log model evidence or marginal likelihood. This leads to a complementary perspective on surprise-minimising dynamics that become self-evidencing; in the sense of optimising Bayesian model evidence – and, by implication, performing some sort of (perceptual) inference. In short, technically speaking, minimising self-information underwrites self-organisation through self-evidencing (Hohwy, 2016); thereby evincing a Bayesian mechanics for any system that exists in the sense of possessing a Markov blanket.

Standard cognitive functions like perception (Hohwy, Roepstorff, & Friston, 2008), attention (Feldman & Friston, 2010), and learning (Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2016a; Friston, 2005) all seem to conform to this single principle. The machinery used to estimate and avoid surprise also recruits a series of non-standard functions like emotions (Van de Cruys & Wagemans, 2011), action (Friston, Mattout, & Kilner, 2011), culture and its production (Fabry, 2017; Ramstead, Veissière, & Kirmayer, 2016), as well as evolutionary processes like niche construction (Constant, Bervoets, Hens, & Van de Cruys, 2018; Constant, Ramstead, et al., 2018) and natural selection (Campbell, 2016; Friston & Stephan, 2007), thereby forcing us to rethink the boundaries of cognition.

Statistically, one can define variational free energy as surprise plus a measure of the distance between a system's (posterior Bayesian) beliefs[13] about the external causes of its sensory input, encoded by its internal states (e.g., neural architecture), and the true posterior probability distribution, conditioned on a generative model of how that input was produced (Friston, 2010). Thus, the variational free energy is defined with reference to a (generative) model of what caused its sensations (including, crucially, its own actions). Variational free

[13] Note that we will use the term 'belief' throughout to mean 'probabilistic Bayesian belief', which is a probability distribution encoded by states of the organism. A belief in this sense need not have any content; it is *not* a belief in the traditional philosophical sense, but instead should be read as synonymous with 'probability distribution'.

energy can thus be cast as a measure of the kinds of things that the cognitive systems finds surprising or, more simply, an estimation of surprise. In summary, variational free energy is an *upper bound* on surprise, in the sense that surprise can never be greater than free energy given the way variational free energy is constructed – for details, see Friston (2012). Thus, by acting to minimise free energy, organisms implicitly minimise surprise.

Crucially, by acting to reduce variational free energy, biological systems come to instantiate a *probabilistic* (*generative*) *model* of their environment, including the states of their body (Friston, Parr, et al., 2017). This generative model can be viewed as a 'map' of the relational or causal structure among the various quantities (e.g., sensory observations and Bayesian beliefs) that are optimized through action, perception, and learning, as the organism navigates, and maintains itself in its environment. Hence, it is said that the generative model is 'entailed' by the existence of an organism (Friston, 2012; Ramstead et al., 2018a), in the sense that it changes as a function of the organism's normal bioregulatory activity. Heuristically, this means that through adaptive action, organisms come to embody a guess about the causes of their sensations (i.e., a generative model) by optimizing its beliefs about those causes.

An intuitive example of free energy bounding dynamics is the maintenance of core body temperature. Human beings tend to maintain their body temperature around 36.5 degrees Celsius. Human bodies expect to be in typical (phenotypical or characteristic) states; surprise is large if the probability of the sensory state is low. So, any deviation from the mean, 36.5 degrees Celsius, implies that the organism is in a sensory state with (relatively) high surprise. Conversely, surprise is low when the probability of the sensory observation is high. Importantly, deviations from the expected (i.e., the mean) state induce *active inference*.

Active inference refers to the joint optimisation of internal states (e.g., perception) and the selection of action policies (i.e., sequences of active states that minimize expected free energy), which function hand-in-hand to reduce free energy (resp. surprise). The system of nested subsystems reacts as a whole, at various scales, to discrepancies between the predictions under the generative model and the actual state of the world. Active inference can take many forms in this setting. Reactions to departures from expected temperature include, at one scale, individual reactions from temperature-sensitive sensory cells in the skin; the raising of individual hairs by skin cells; the registering of a temperature difference by the networks of the nervous system, and the body's subsequent engaging in shivering behaviour. More individual, psychological reactions to changes in temperature might include enjoying this change (or not); culturally-mediated behavioural reactions to differences in temperature

might come into play as well, relying on elements of the cultural niche. If it is too hot, we might take off some clothes; but if one lives in the desert, this exposes one's bare skin to the elements; and to fend off the heat, we might instead put on robes, as Bedouins do in the desert.

2.2. Generative models and action policies

In the variational approach, the form taken by the generative models is that of *graphical models* (Friston, Parr, et al., 2017). The model itself carries *correlational information about causal factors that lead to the generation of sensory states*. So, in a nutshell, the model is *intrinsically* probabilistic and correlational, not causal; in the sense that the generative model, by necessity, captures useful probabilistic information about the agent acting in its niche.

Technically, the generative model is just a probability distribution over the joint occurrence of sensory states (of the Markov blanket) and the external states generating sensory states. It is a *normative* model, in the sense that it specifies the conditions that allow the continued existence of the type of creature being considered. This can be variously formulated in terms of the likelihood of some sensory states, given external states and prior beliefs over external states. It manifests in active inference via inferential dynamics (i.e., action and perception) that flow on free energy gradients, where the free energy is defined in terms of a generative model.

However, the variational story is one about how the respective statistical structures of the generative model and generative process (the actual causal structure that generated observations) become attuned to one another. So, when everything is going well (i.e., when the organism engages in adaptive behaviour and thrives in its niche), the correlational structure carried by the generative model – ideally – maps onto the causal structure of the generative process in the environment. So, while the model is necessarily only ever probabilistic, it remains that active inference fits or tunes the generative model to the generative process; and by that fact, the generative model gains some causal purchase: indeed, the generative model is often described as a probabilistic description of how sensory consequences are generated from their causes. Inference then corresponds to the inversion of this mapping – to infer causes from consequences. This is inference is, by construction, implicit in the minimisation of free energy or the maximisation of model evidence.

One novel way to think about the generative model is in terms of 'enactment'. On this view, minimising free energy essentially means reducing the disattunement between the expectations of an organism and the generative model under which actions are selected

(Bruineberg & Rietveld, 2014). Active inference is the process of creating and maintaining self-organization through action. Under the FEP, active sampling of sensory states is a feature of the entire dynamics themselves, which entail a generative model. This speaks to the idea that the entire process of attuning the system to its niche involves perceptual inference, but especially the selection and expression of relevant *action policies* – policies that select the actions most likely to elude surprise. Minimising expected surprise does not mean avoiding sensations, on the contrary, it means resolving uncertainty by seeking out salient, informative sensations. This follows simply from the fact that *expected surprise* (i.e., self-information) corresponds to *uncertainty* (i.e., entropy) (Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2016b; Friston, Rosch, Parr, Price, & Bowman, 2018)

This implies that the function of the generative model is to guide action in a context-sensitive fashion; in turn, this speaks to a shift away from viewing the brain in terms of Bayesian predictive processing to how the brain enables "feedback loops that maintain attunement to the environment and support adaptive behavior" (Anderson, 2017, p. 8). This dynamic emphasis on the realisation of biological self-organisation through adaptive action clearly aligns the FEP with enactive and pragmatist approaches to cognition (Bruineberg, Kiverstein, & Rietveld, 2016; Engel, Friston, & Kragic, 2016; Kirchhoff & Froese, 2017; Ramstead et al., 2019) – a point we will explore in greater detail in section 4.

2.3. Markov blankets and the boundaries of cognitive systems

Under the FEP, the statistical conception of life leads to a formal, statistical ontology of living systems (Friston, 2013; Kirchhoff et al., 2018; Ramstead et al., 2018a). This ontology leverages a statistical formalism; namely, the Markov blanket formalism, which provides a principled account of what constitutes a system, and what does not. A Markov blanket is a statistical partitioning of a system into internal states and external (i.e., non-constitutive) states; where the blanket itself can be partitioned further into sensory and active states (Clark, 2017; Friston, 2013; Pearl, 1988). This implies that internal and external states are conditionally independent from one another, given that internal and external states can only influence each other via sensory and active states.

A Markov blanket constitutes the evidential or existential boundary that sets something apart from that which it is not. A cell therefore has a Markov blanket – its plasmalemma. As do multicellular organisms like *Homo Sapiens*. Take the cell as an example. It arises out of a molecular soup by assembling its own boundaries, thus acquiring an identity (Friston, 2013; Varela et al., 1991). For a cell to remain alive, its internal states

must constantly organise and prepare its boundaries – lest it decay, and dissipate into its surroundings (Di Paolo, 2009).

This, in turn, implies the maintenance of a statistical boundary that separates internal from external states, and *vice versa* (Friston, 2013). Under the FEP, this statistical boundary is an *achievement*, rather than a given; it is generated and maintained through active inference (i.e., adaptive action). This again aligns the FEP with enactive and pragmatist approaches to cognition (Engel et al., 2016). Thus, under the FEP, to exist 'just is' maintaining the states that comprises one's Markov blanket through active inference. In other words, without a Markov blanket and the processes that assemble it, the cell would cease to exist, as there would be no way for the cell to restrict itself to a characteristic set of states. In other words, there would be no way of establishing the conditional independence between internal states and the surrounding environment – and the cell would simply dissolve, dissipate or decay into its universe (Hohwy, 2016).

The nice thing about Markov blankets is that they allow us to speak in a meaningful (and mathematically tractable) way about conditional independencies between internal and external states. Consider again the cell. The intracellular (i.e., internal) states of a cell have an existence that is distinct from their external environment. This shows that intracellular and extracellular states are conditionally independent. It is the conditional independence (in a statistical sense) between internal and external states that are captured – or indeed defined – by appeal to the concept of a Markov blanket (see Figure 1).

2.4. A formal ontology for the boundaries of cognitive systems

This reading of active inference as self-evidencing makes the boundary of cognitive systems an *existential* notion, tied up with the epistemic process of generating evidence for your own existence. In a nutshell, then, to enact a generative model is to provide evidence (i.e., to generate evidence through adaptive action) for a model of one's existence.

More specifically, the claim we are making about the status of the boundary of cognitive systems is that *this boundary is both ontological and epistemological*. The boundary of a given cognitive system is given by the Markov blanket of that system, which carves out or individuates a system by separating systemic states from non-systemic ones. The Markov blanket is an *ontological* boundary, in the sense that this boundary individuates the system as the kind of system that it is. It sets apart the states that count as systemic states from those that count as part of its surroundings. Markov blankets provide the most minimalistic answer to this question, based on the notion of *conditional independence*. If a

system exists, there must a sense in which the non-systemic parts can change without the system of interest changing in concert. Markov blankets formalise this requirement. The Markov blanket is a result of the system's dynamics (i.e., the system's patterns of adaptive action), which means that it is the system's dynamics itself that carves out the relevant boundaries. In other words, the boundary itself is orchestrated and maintained through active inference, it is an achievement of the cognitive system that is orchestrated and maintained through adaptive action.

We claim that the Markov blanket is an *epistemological* boundary as well. This is because the boundary is realised through active inference, which is a process of *self-evidencing*. Self-evidencing means that *to exist as a system* is to produce *evidence of your existence*. More explicitly, the variational framework suggests that the dynamics of living systems entails a *generative model* of one's existence. The variational framework tells us how the generative model that organisms embody and enact tunes itself to (approximates the statistical structure of) the generative process, or actual causal process in the environment that causes the sensory states of an organism. To exist as a living being and to engage in adaptive action (when all goes well) *just is* to realise the relations between quantities that are modelled in the generative model. In other words, under the FEP, to exist at all means to produce evidence for a model of oneself (or more exactly, since the generative model is a control system, a model of oneself acting in the world).[14] Existence in this sense is fundamentally tied up with the creation and maintenance of an informational boundary, i.e. the Markov blanket.

The Markov blanket formalism, then, tells us what counts as a system and what does not. It provides us with a principled means to determine what it is to be a *self-evidencing system* under the FEP. In this sense, the term *existential* boundary might be most appropriate: the evidential boundary is also an existential boundary.

In summary, when applied to the biological realm, the statistical formalism of the Markov blanket provides a way to define the boundaries of a system. To so enshroud the internal (constitutive or insular) states of a system behind a Markov blanket enables the individuation of a well-defined partition of the system into internal and external states,

---

[14] A very similar conflation of epistemological and ontological notions of a 'model' was apparent at the inception of cybernetics in the form of the Good Regulator Theorem (every system that regulates its environment must be a good model of that environment) (Conant & Ashby, 1970). The free energy principle formalises this notion by equipping existential dynamics with an epistemological corollary cast in terms of inference.

mediated by the (active and sensory) states that comprise the Markov blanket itself, and over which we can define systemic dynamics.

3. Cognitive boundaries: Externalism and internalism

In this section we have two agendas. The first is to address externalist or radical views of cognition; namely, embodied, enactive, and extended cognition. We will pay special attention to enactive formulations of life and mind, highlighting that on this account, the basis of life and mind is a nested set of properties: autopoiesis, operational closure, autonomy, and adaptivity. The nice thing about this formulation of living and cognitive systems is that it allows us both to address the organisational principles of life, as per the enactive framework, and speak to how this framework underpins the ideas of cognition as realised across brain, body, and world; while, at the same time, giving a special place to embodied activity in the assembly of cognitive activities and processes. Our second agenda is to describe how this emphasis on (especially) adaptive operational closure could be turned into an argument against the enactive view by appeal to the active inference scheme and the Markov blanket formalism.

3.1. Externalism: Radical views of cognition

Embodied approaches to cognition hold that the body is crucial for cognition (Gallagher, 2006). Extended views suggest that not only are bodies important, the local environment of individual cognitive systems can partly realise cognitive processes (Clark, 2008; Clark & Chalmers, 1998). Enactive views play up the role of action in the functioning of cognition, especially on certain accounts of enactivism tethering mind to the biology of living systems (Chemero, 2009; Gallagher, 2017; Thompson, 2010). In this subsection we formulate the enactive view associated with the work of Varela and colleagues; so-called autopoietic enactivism (Di Paolo, 2009; Di Paolo & Thompson, 2014; Thompson, 2010; Varela et al., 1991). Our focus is selective; the enactive framework not only exemplifies current radical views on cognition, it also shares a number of important overlaps with our multiscale integrationist view, derived from the FEP.

A central aspect of living and cognitive systems is their *individuation*. Individuation is the process that makes something distinct from something else, and is in this sense consistent with our use of the Markov blanket formalism as a means by which to delineate systemic boundaries separating systemic from non-systemic states, and *vice-versa*. Crucially, on the enactive account, this process of individuation implies that systems that can self-organise

their own process of individuation are (a) *autopoietic*, (b) *operationally closed*, and (c) *autonomous*. Autopoiesis denotes the property of structural self-generation; namely, the capacities to (re-)generate and maintain systemic constituents, despite compositional and functional change. An autopoietic system can be cast as an operationally closed system. Operational closure refers to processes of autopoietic self-assembly, on the one hand, and boundary conservation conditioned on interdependent processes, on the other. This is entirely consistent with the kind of statistical independence between states induced by the Markov blanket formalism, as this implies that the very existence of a living system is premised on recurrent processes that work to conserve the integrity of systemic boundaries (see Figure 3).

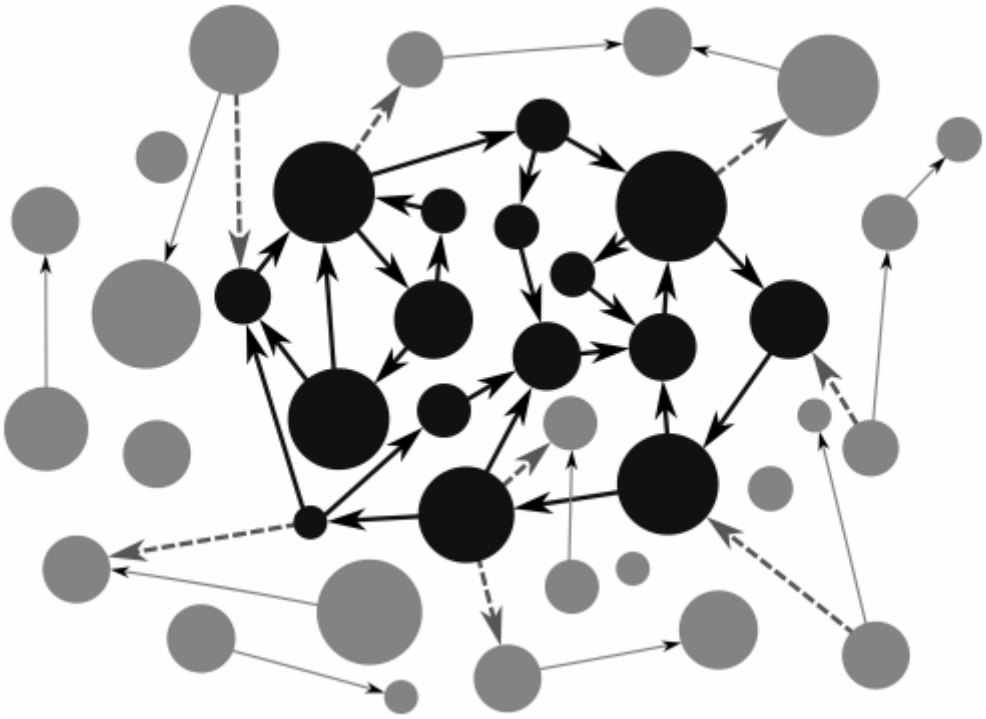


Fig. 3. An illustration of operational closure. Here the black circles form part of an operationally closed network of self-organising processes. Each black circle has at least one arrow arriving at it and at least one arrow coming from it – respectively originating and ending in another black circle. Dashed arrows refer to enabling relations between processes in the operationally closed network and processes that do not belong to it. Adapted from Di Paolo and Thompson (2014, p. 70).

In an operationally closed network each process is affected by another process such that the operations of processes comprising the network are dependent on each other. As Di Paolo

and Thompson put it, in relation to this figure: “If we look at any process in black, we observe that it has some enabling arrows arriving at it that originate in other processes in black, and moreover, that it has some enabling arrows coming out of it that end up also in other processes in black. When this condition is met, the black processes form a network of enabling relations; this network property is what we mean by operational closure.” Di Paolo and Thompson (2014, p. 71). To make this a little more concrete, consider Figure 4.

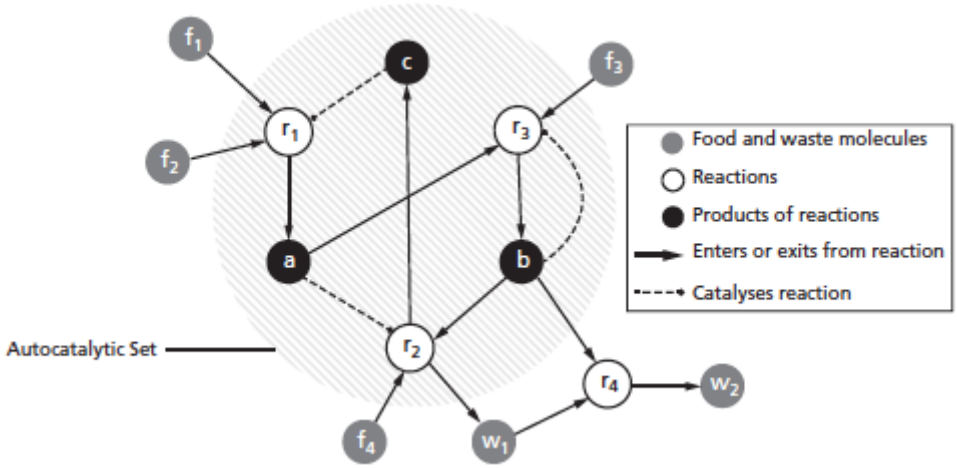


Fig. 4. Schematic illustration of autocatalytic closure. A network of chemical reactions is organized such that each reaction is enabled or catalyzed by products of other reactions in the network. From Di Paolo, Buhrmann, and Barandiaran (2017, p. 113).

This figure describes a network of four reactions, r_1, r_2, r_3, and r_4, each of which is enabled – in the sense of being accelerated to sufficiently fast rates – by the molecules of type, *a*, *b*, and *c*, which are themselves the products of the same reactions (Di Paolo et al., 2017). This is an example of an operationally closed network, given that – as a whole – the set is able to enable its own production.

Given what we have said in Section 2, it is fairly straightforward to establish that the Markov blanket formalism provides a statistical formulation of operational closure (Kirchhoff et al., 2018). In the same way that active and sensory states of Markov blankets couple internal and external states – via an informational dynamics – operational closure does not imply that the systemic (i.e., operationally enclosed) states are cut-off from external states. To see this, note that autonomy implies that an operationally closed network of self-enabling processes can modulate its relation to the embedding environment. If this were not so, the network would stop or run down. The nice thing about the emphasis on autonomy is that it

speaks directly to *adaptivity*, the basic capacity to act purposefully and probabilistically, as the basis of the self-organisation of life and cognition (Di Paolo, 2005). In the context of the FEP, this is called *adaptive active inference* (Kirchhoff, 2018a, 2018b).

This enactive view of living and cognitive systems exemplifies a radical view of cognition; i.e., a view that breaks faith with the standard assumptions about internalism. First, autopoietic enactivism is a denial of any kind of internalism, given that it is entirely possible for operationally closed dynamics to be realised in an extensive network of processes breaking across neural and non-neural variables (De Jaegher & Di Paolo, 2007). Second, autopoietic enactivism denies what is a usual starting point of so-called first-wave or functionalist arguments for the extended mind thesis. First-wave arguments starts by taking the individual as the default cogniser and only then asks whether some worldly elements can play functionally similar roles to mental states or cognitive processes realised internally (Clark & Chalmers, 1998). So these arguments for the extended mind assume a kind of internalism in their formulation (for similar critiques, see Kirchhoff, 2012; Menary, 2010). Finally, autopoietic enactivism holds the view that cognition is a relational phenomenon between an organism and its environment (Thompson & Stapleton, 2009).

3.2. Internalism: Pushing back

Despite its influence in the sciences of life and mind, the enactive approach can be put under pressure. Indeed, a specific formulation of the FEP, turning on the Markov blanket formalism, arguably pushed back against any of these radical views of cognition (Hohwy, 2016, 2017). Our aim in this subsection is to (briefly) rehearse some of the main steps of this internalist argument. We develop a counterargument to the internalist position in the next section, which gives an information-theoretic justification for the view that the boundaries of cognition are nested and multiple.

In a nutshell, the internalist argument states that the relevant boundary for cognitive systems and cognition is essentially the boundary of the brain or skull. Internalists take the inferential seclusion of internal states in active inference – that internal (systemic) states are hidden behind the veil of the statistical Markov blanket – to imply that the boundaries of cognition stop at the boundaries of the brain given the presence of a brain-bound Markov blanket (Hohwy, 2016; Seth, 2014).

A crucial aspect of this argument is the assumption that the brain itself is a generative model of its environment, one that "garners evidence for itself by explaining away sensory

input" (Hohwy, 2017, p. 1) by a process of variational Bayesian inference[15]. This means that through active inference, a cognitive system minimises its variational free energy, thereby securing the evidence for its generative model, and inferring the hidden causes of its observations (sensory data). Cognitive processes (e.g., attention, learning, decision, perception, and so on) are processes that work to optimise internal states in accordance with the FEP – and implicit self-evidencing (Hohwy, 2016).

The next, crucial step is the statistical partitioning of a system into internal and external states through the Markov blanket formalism. This captures the notion that internal (neural) and external (environmental) states are conditionally independent; capable of influencing one another only via sensory and active states. Internalists interpret the Markov blanket as enforcing an evidentiary boundary severing, in an epistemic and causal sense, the brain from its body and environment. Thus, Hohwy concludes: the "mind begins where sensory input is delivered through exteroceptive, proprioceptive and interoceptive receptors and ends where proprioceptive predictions are delivered, mainly in the spinal cord." (Hohwy, 2016, p. 276).

The issue of *internalism* comes up when these two notions, the Markov blanket and active inference, are combined in the free energy formulation. Proponents of internalist readings of the FEP argue that the presence of a Markov blanket implies that systems that minimise their free energy, on average and over time, ipso facto, are epistemically and causally secluded from their environment. The upshot of such a conception is that the boundaries of cognition stop at the skull. Mind is a skull-bound phenomenon. The rationale for this way of thinking is that the spontaneous formation of Markov ensembles realizes a form of Bayesian inference (active inference). Active inference carves out coherent neural ensembles, which are neural ensembles (Hebbian assemblies) 'wrapped' in a Markov blanket (Hohwy, 2016; Yufik & Friston, 2016). This means that cognition implies the transient assembly of such brain-bound Markovian ensembles.

This internalist rendition of internal states, hidden behind of curtains of the Markov blanket, leads to a neo-Kantian or Helmholtzian account of cognition that emphasises its indirect nature (Anderson, 2017; Bruineberg et al., 2016). The Markov ensembles are said to infer external states, and this inference is taken to be a content-involving affair. This means

[15]Also known as approximate Bayesian inference. Variational or approximate Bayesian inference necessarily entails the minimisation of variational free energy, under a generative model and an assumed form for posterior beliefs.

that inferences are over content-involving states (in the sense that internal states that are *about* things in the world), which are cast as hypotheses and beliefs[16]. The idea is that organisms leverage their generative model to infer the most likely hidden causes of its sensory states. This is a Helmholtzian interpretation of the FEP (Bruineberg et al., 2016). On this reading, active inference is understood on an analogy with scientific inference, as literal hypothesis-testing.

This kind of neo-Kantian schism between mind and world is taken to imply that the contact of a cognitive system with its environment – perceptually or behaviourally – is mediated by its internal (neural) states, often interpreted as representations encoded in hierarchical generative models that are realised in the brain's cortical architecture (Gładziejewski, 2016; Gładziejewski & Miłkowski, 2017; Williams, 2017). We shall not dwell on the question whether these internal states are representations – for contrasting interpretations, see (Kirchhoff & Robertson, 2018) versus (Kiefer & Hohwy, 2017). However, in the next section, we consider whether internalist interpretations of Markov blankets and generative models are appropriate.

4. Multiscale Integration: Nested and multiple boundaries

In this section, we argue that the internalist interpretation of the boundaries of cognition rests on a problematic interpretation of what generative models are, and the kind of properties they have under the FEP. Crucially, we agree with internalism that any relevant mind-world relation is mediated by processes that can be cast as both assembling and finessing the generative model. But this is just to say that we can describe how internal and external states are statistically coupled to one another via intricate and complex sensorimotor dynamics (Gallagher & Allen, 2016; Kirchhoff & Froese, 2017).

4.1. Generative models: what they are, and how they are used to study cognition

First, we take issue with the claim that, under the FEP, the generative model is something internal to the organism (i.e., that the generative model comprises neuronal vehicles or any other vehicles). Rather, the generative model is a mathematical construct that explains how the quantities embodied by the system's architecture change to transcribe (i.e., update beliefs

[16] This 'aboutness' is a key aspect of the FEP and follows from the fact that internal states encode or parameterise a probability distribution over external states. In other words, internal states are the sufficient statistics of Bayesian beliefs (about external states). See Ramstead et al. (2018a, Box 4).

about) the causes of the system's sensory observation. What should be at stake in the debate between internalists and externalists is the status of the 'guess' that the organism embodies; namely, the *posterior beliefs*[17] encoded by internal states, and whether this guess does, or does not constitute the limit of 'cognition', understood as the avoidance of surprisal, or informational homeostasis.

The *posterior belief* (i.e., recognition density) represents the system's 'best guess' regarding the causes of its sensory states, and is embodied or encoded by the states of the organism; technically, internal states of the Markov blanket (Friston, 2012). Under the FEP, the *system's* posterior belief is refined or 'tuned' under the generative model, through a process of variational (approximate Bayesian) inference, and becomes a tight bound on the *true* posterior belief it aspires to (Friston et al., 2016b).

The generative model is a statistical construct that transcribes the expected sensory causal regularities in the process generating sensory states. The generative model is used to model the set of viable phenotypical statistical relations (preferences, and action policies) that must be brought forth by the organism in active inference: in short, a model of a viable state of being for the organism. Through active inference, internal states are tuned and this tuning changes its posterior belief, and hence the organism's 'best guess' about what caused its sensations (that usually include its own actions). In other words, a generative model can be used to understand how organisms are able to track (infer) their own behaviour.

The FEP is based on the idea that the functions and changes in the structure of living systems conform to approximate Bayesian inference. This assumption rests on the claim according to which living systems avoid surprise (cf. Section 2.); approximate Bayesian inference under the FEP, then, is just one sensible strategy to understand how living systems avoid surprise and the dispersion of their sensory states. It rests on what we described earlier as the generative model (the control system), the recognition density (the living system), and the generative process (the external world, which includes the organism's actions). Simply put, the relation between these is that the recognition density changes as a function of the control system; and because the control system constitutes expectations about the world conditioned upon the preferences of the living system, the living system turns out to change

[17] By definition, the posterior belief encoded by internal states approximates the true posterior belief when free energy is minimised. It is often referred to as a recognition density or approximate posterior belief.

so as to become (statistically) consistent with the preferred world; that is, according to its preferences and expectations about the world.

Under the FEP, 'cognition' is what the recognition density, or living system does (i.e., changing to elude surprises and maintaining informational homeostasis by minimizing free energy), and the way one studies cognition (i.e., what the system does), is by developing, simulating, and analysing the possible generative models that explain how the recognition density of interest (the system of interest) changes so as to attain minimal free energy.

In other words, 'drawing the bounds cognition' means defining the *recognition density* of the system of interest, and identifying a *generative* model that explains changes in that system that follow variational Bayesian inference[18]. In this sense, cognitive science might be understood as the study of generative models and processes: it is in the business of modelling the correlational or causal structure of actions and observations of the organism. The generative model, then, is not the *vehicle* of something like content or mutual information; instead, it is the tool that we use to study cognitive systems (as explanatory model), and indeed, perhaps more speculatively, the guide, or path living systems entail and follow to stay alive (as control systems). The vehicle is the recognition density (also called the variational density), the 'best guess' that the system of interest embodies, and whose function and structure can be studied using the generative model.

This means that we can study cognition meaningfully as it occurs in individuated systems at the respective scales at which those systems exist; e.g., the brain in ontogeny, or large-scale ensembles like species over phylogenetic time. Since organization at each level depends upon the integration to the entire dynamics, one can also study cognition 'across boundaries'. Below, we will see that we can formalise how the system moves from one state to the other in terms of a free energy bounding dynamics. This dynamics integrates systems of systems; all individuated as nested Markov blankets of Markov blankets.

In summary, we are suggesting that organisms use a statistical trick – i.e., the minimization of variational free energy – to track the causes of their sensory states and to select appropriate actions. The key is to note that organisms are organized such that they instantiate the prior that their actions will minimise free energy. This mechanics of belief is

---

[18] This is especially true in specialised fields such as computational psychiatry, where a crucial part of the generative model – namely prior beliefs – completely specify behaviours and preferences. This means that any behavioural phenotype can be, in principle, quantified or understood in terms of prior beliefs. See Schwartenbeck and Friston (2016) for a worked example.

the only causally relevant aspect of the variational free energy. The free energy may or may not exist; what is at stake is the causal consequences of the *action-guiding beliefs of organisms and groups of organisms*, which are harnessed and finessed in the *generative model* (Ramstead et al., 2019). What matters is that organisms are organized such that they instantiate such a prior to guide their action.

4.2. Enactivism 2.0.

The generative model, as we have seen, functions as a *control system*. That is, its function for the cognitive system is to generate of adaptive patterns of behaviour. In the parlance of the FEP, its purpose is to guide the evaluation and selection of *relevant action policies* (Friston et al., 2016b). The generative model is a strange beast in the variational framework, in that it exists only insofar as it underwrites the organism's inference about states of affairs and subsequent action selection. Since the free energy expected following an action, which determines the policy to be selected, is defined in terms of the generative model, the latter is the cornerstone of the self-evidencing process.

This emphasis on adaptive action aligns active inference with one brand of radical accounts of cognition, namely *enactivism*. Indeed, it has been argued that the FEP provides an implementation of enactivism, and in a sense supersedes or absorbs classical (i.e., autopoietic) formulations of enactivism (i.e., Froese & Di Paolo, 2011; Thompson, 2010). (See Kirchhoff, 2018a; Kirchhoff & Froese, 2017, for a detailed argument to this effect). Active inference is inherently a pragmatist or enactive formulation, and can be contrasted with non-enactive appeals to Bayesian principles of cognition, such as predictive coding.

However, because it relies fundamentally on formulations from information theory, active inference is in tension with a few of the more (arguably) conservative elements of the enactive theory. Indeed, classical enactivism has typically rejected the appeal to information theory to describe cognition (e.g., Thompson, 2010). We believe this is a hangover from another age in cognitive science. And, more to the point, this conservatism has not prevented the proponents of active inference from taking up the banner of enactivism (Bruineberg et al., 2016; Engel et al., 2016; Kirchhoff & Robertson, 2018; Ramstead et al., 2018a). Active inference provides a theoretical model for enactment. Allen (2018) has called this form of enactivism, based in information theory, 'enactivism 2.0', or Bayesian enactivism.

4.3. Nestedness: or how to study cognition beyond the brain

The existence of Markov blankets at one scale means that interaction amongst components at that scale are mediated by states belonging to their respective Markov blankets. These active exchanges have a sparsity structure that induces *nested sets of Markov blankets* – that is, Markov blankets of Markov blankets (Kirchhoff et al., 2018; Ramstead et al., 2018a). The central idea behind the multiscale integration of Markov blankets is that the particular statistical form and the specific partitioning rule that governs the Markov blanket allows for the assembly for larger and larger Markov blankets (of cells, of organs, of organisms, of environments, and so on). This is because Markov blankets at increasingly larger scales of systemic organisation recapitulate the statistical form of Markov blankets at smaller microscopic scales of systemic organisation. This can be shown to follow from the observation that any meaningful statistical separation between internal and external states at the scale of, for example, complex organisms, a macroscale Markov blanket must be present, whose sensory and active states, distinguish this organism from its local niche, and which itself is composed of smaller and smaller Markov blankets sharing the same statistical form as the macroscopic Markov blanket (see Figure 5)

Figure 5 illustrates the idea of Markov blanket formation at any scale of hierarchical and systemic organisation, thus speaking to the notion that organisms and their local environment will be "defined not by a singular Markov blanket, but by a near-infinite regress of causally interacting Markov blankets within Markov blankets." (Allen & Friston, 2016, p. 19). This, in turn, provides an integrated perspective from which to approach the multiple scales of self-organisation in living systems.

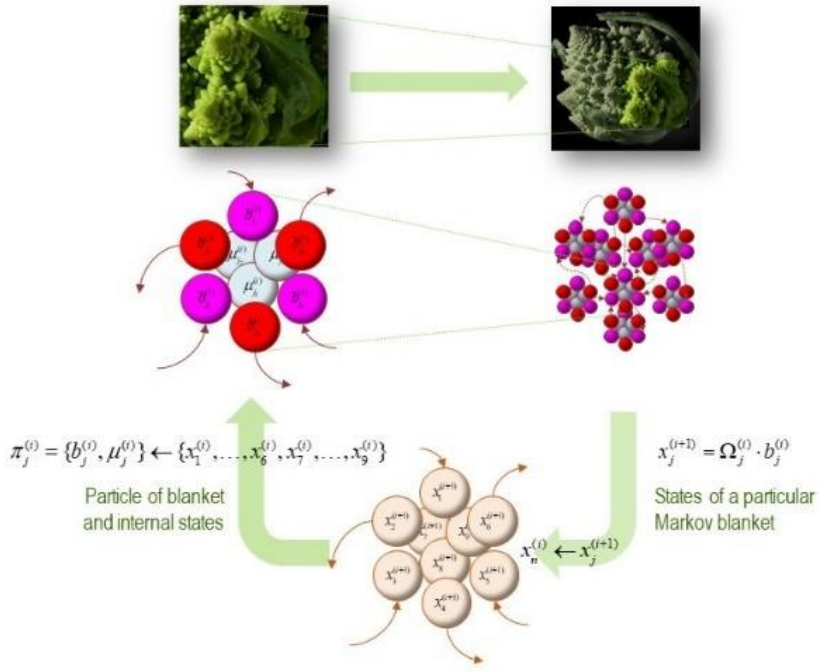


Fig. 5. *Blankets of blankets*. This figure depicts the recursively nested structure of Markov blankets that forms the basis of our formal ontology of cognitive systems. In this scheme, successively larger and slower scale dynamics arise from, and constrain, those dynamics arising at subordinate (smaller and faster) scales. Consider an ensemble of vector states (here, in the lower panel, nine such states are depicted). The conditional dependencies between these vector states define a particular partition of the system into *particles* (upper panels). The effect of this partition into particles is, in turn, to partition each of these particles into blanket states and internal states. Blanket states comprise active states (red) and sensory states (magenta). Given this new partition, we can summarize the behaviour of each particle in terms of (slow) eigenmodes or mixtures of its blanket states, which in turn produces vector states at the next (higher) scale. These constitute an ensemble of vector states and the process can begin anew. The upper panels depict this bipartition into active and sensory states for a single particle (left panel) and for an ensemble of particles. The insets at the top of the figure illustrate the self-similarity that arises as we move from one scale to the next. In this figure, $\Omega \cdot b$ denotes a linear mixture of blanket states that decay sufficiently slowly to contribute to the dynamics at the level above. Adapted from Ramstead et al. (2018a).

The multiscale partition of model parameters, encoded by internal states of the Markov blanket, attunes itself to the sufficient statistics of the generative process that generated the sensory observations, tuning its internal states by bounding free energy. This process occurs at and across spatiotemporal scales, effectively integrating the system through dynamics. Indeed, for each system individuated at a given scale, one can define a generative

model entailed by the dynamics at the scale above; which speaks to the complementarity between specialisation and statistical segregation, on the other hand, and functional integration, on the other (Badcock, Friston, & Ramstead, 2019).

Free energy is an additive or extensive quantity minimised by a multiscale dynamics integrating the entire system across its spatiotemporal partitions (Ramstead et al., 2019). There is also, therefore, only one free energy for the entire system, which is simply the sum of free energies at all the relevant scales (see Figure 6). The whole system dynamics leverage internal states across temporal and spatial scales, to integrate the system across scales. This means that the variational approach accommodates both multiscale partition of the recognition density, and a multiscale integration (through active inference).

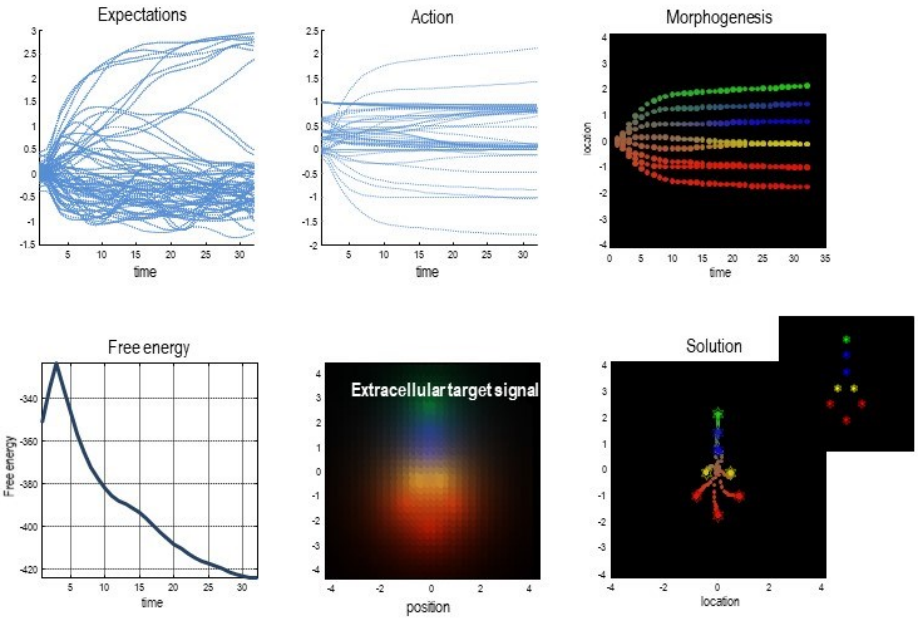


Fig. 6. *Multiscale self-organization and active inference*. This figure depicts variational free energy being minimised across scales through active inference. It presents the results from a simulation of morphogenesis using the active inference framework (Friston et al., 2015). The simulation used a gradient descent on variational free energy to simulate a group of cells self-assembling into a larger pattern (i.e., target morphology). The simulation employed an ensemble of eight cells. Each cell was equipped with the same generative model,

which is a metaphor for shared genetic information. This generative model generated a prediction of what each cell would sense and signal to other cells (chemotactically) for any given location in a target morphology. In other words, the model predicted what each cell would expect to sense and signal if it were in that location (lower middle panel – extracellular target signal). Each cell engaged in active inference, by actively moving around to infer its place in the target morphology relative to other cells. In doing so, each cell minimised its own variational free energy (and by proxy, its surprise or self-information). Remarkably, the fact that all cells shared the same generative model allowed their individual active inference to minimise the free energy of the ensemble, which exists at the scale above the individual cells. Each of the cells that make up the ensemble shares the same generative model. Crucially, the sensory evidence for the model with which each cell is equipped is generated by another cell. The arrangement that minimises the free energy of the ensemble is the target morphology. This means that each cell has to 'find its place'; the configuration in which they all have found their place is the one where each cell minimises its own surprise about the signals it senses (because it knows its place), and in which the ensemble minimises the total free energy as well. The upper panels show the time courses of expectations about the place of each cell in the target morphology (upper left), as well as the associated active states that mediate cell migration and signal expression (upper middle). The resulting trajectories have been projected onto the first (vertical) direction and color-coded to show cell differentiation (upper right). The trajectories of each individual cell progressively, and collectively, minimize total free energy of the entire ensemble (lower left panel), which illustrates the minimization of free energy across scales. The lower right panel shows the configuration that results from active inference. Here, the trajectory is shown in small circles (for each time step). The insert corresponds to the target configuration. In short, all multiscale ensembles that are able to endure over time must destroy free energy gradients, which integrates system dynamics within and between scales. Adapted from Friston et al. (2015).

The underlying philosophical point is that states that are statistically isolated by Markov boundaries become integrated under one dynamics in active inference; they come to parameterise one generative model (the one entailed by the adaptive behaviour of the whole system), thereby guiding one integrated action across temporal and spatial scales. Internal states that are inferentially secluded at one scale become absorbed into higher order Markov blankets and dynamics at the scale above. This means that the epistemic seclusion of internalism is, in a sense, illusory or partial; since the entire organism engages in active inference across scales. Under the FEP, inferential seclusion coexists with existential (pragmatic) integration through dynamics (i.e., adaptive behaviour).

This perspective vindicates an integrationist ontology of the boundaries of cognition, while retaining the possibility of granting epistemic priority to any of these boundaries, given explanatory interests. The nested Markov blankets perspective answers the question of how to understand the generative model from this multiscale perspective. The challenge, now, is

to develop the theoretical apparatus to describe the boundaries of cognition at higher scales. We address this issue in the next subsection.

In a nutshell, at any scale, the relevant Markov-blanketed systems are composed of parts that, in virtue of their (relative) conditional independence, can also be described as Markov blanketed systems. Each of these separate Markov blanketed subsystems might count as separate systems, i.e., one cognitive subsystem can be a nested part of another system. However, all these nested boundaries are integrated within the same system. More precisely, all the subsystems that are individuated by their own Markov blanket are *integrated* as one single dynamical system through the *system dynamics* (i.e., *adaptive action*). Collectively, there is only one (hierarchical) generative model and therefore one free energy functional, for the ensemble of nested blankets (where each constituent blanket itself has a generative model and accompanying free energy functional). This sort of nesting is particularly prescient for hierarchical systems like the brain. In this brain-bound setting, the integrated Markov blanket could be regarded as comprising the brain's sensory epithelia and motor (or autonomic) efferents, while internally nested Markov blankets are a necessary feature of neuronal (e.g., cortical) hierarchies (Shipp, 2016; Zeki, 2005; Zeki & Shipp, 1988). At each and every level of the cortical hierarchy the associated free energy is minimised by neuronal dynamics, such that the total free energy of the brain is upper-bounded in accord with the FEP.

4.4. Multiplicity: or how to describe cognition beyond the brain

Central to our discussion is the concept of joint phenotype, which we have introduced in Section 1 in terms of repertoire of highly probable states and traits. Some of those states are contained within the organism (e.g., brain states), and other traits extend far beyond the internal states of an organism (e.g., states of the niche). We use the concept of joint phenotype to support our description of the boundaries of cognition at higher scales.

Typically, joint phenotypes are seen as shared 'extended phenotypes' (Dawkins, 1982). Extended phenotypes are traits (e.g., niche construction outcomes like beaver dams) that, like physiological states, undergo selection due to their fitness enhancing impact. In the case of an extended trait, the impact is on the genes having favoured the reproduction of that extended trait (e.g., beavers' genetic disposition to build dams).

Extended phenotypes, therefore, are extensions from genes to the extended trait. Accordingly, the typical view of the joint phenotype broadly construed describes coextensive phenotypic traits consistent with two or more different species' genetic makeup. In that case,

all parties can be 'joint owners' of the trait; for instance, the insect and the plant are joint owners of *the portion of the leaf* eaten by the insect (Queller, 2014).

The FEP interpretation of the joint phenotype that interests us here brings this a step further. On that view, coextensive phenotypic traits *do not* need to be included in the extended phenotype. They can include biotic or abiotic traits, like ecological cascades produced by niche construction, or other 'seemingly' random effects of organismic activity. These are not directly related to the genetic makeup of either party, while nonetheless being seen as having a systematic and evolutionary significant impact on fitness.

With the FEP, one can study organism-niche complementarity that obtains through phenotypic accommodation and niche construction over development (i.e., adaptation) using variational free energy (Bruineberg, Rietveld, Parr, van Maanen, & Friston, 2018; Constant, Ramstead, et al., 2018), and thereby predict the influence of a trait on fitness. Hence, one can conceive of and study joint phenotypic traits as non-genetically specified traits by studying the changes in the statistical relationship that bounds those traits to the states of the organism(s).

Now, the point we want to motivate here is that – especially in humans – many traits of the constructed niche defining the human joint phenotype increase state-trait complementarity by smoothing the attunement process, or variational free energy minimising process. For instance, in developmental psychology and niche construction theory, it is argued that the material artefacts populating human niches enable individuals to deal with perceptual uncertainty (Christopoulos & Tobler, 2016; Dissanayake, 2009) by constraining, and directing sensory fluctuations in their surrounding (Constant, Bervoets, et al., 2018).

Briefly, computing expected free energy requires computing the cost of a policy (where the cost is given in terms of the divergence between posterior beliefs and preference about sensory outcomes), and the expected ambiguity, or expected 'certainty' about the sensory outcome relative to one's beliefs about the state of the world (i.e., expected surprise) (see Friston et al. 2016a,b for a detailed treatment). Artefacts that populate human niches can be seen as doing much of the legwork in computing the expected ambiguity term that constitute expected free energy. In that sense, they ease the modelling activity of the organism, understood as expected variational free energy minimization (Constant, Ramstead, et al., 2018); cf. epistemic affordance (Parr & Friston, 2017).

Thus, especially in humans, when taking the FEP perspective, one can include external, joint phenotypic traits within the boundaries of cognition for higher scales systems like joint phenotypes, or bodies-environment systems. It also means that under the FEP, one

could meaningfully study cognition 'from outside the brain', for instance, by producing a generative model of an higher scale system (e.g., that of the leaf-insect system) and by simulating the effects of external factors on variational free energy, like environmental cues (Sutton, 2007); cultural practices (Vygotsky, 1978); ecological information (Gibson, 1979). Again, this speaks to the idea that the relevant boundaries of cognitive systems are relative to explanatory interests (e.g., cognition from the point of view of neurophysiology for cognitive neuroscientists, or cognition from the point of view of ecology, for behavioural ecologists).

The Markov blanket formalism might allow us to study the transient assembly of cognitive boundaries over time, in the spirit of the models considered above. Indeed, the original simulation studies employing the Markov blanket formalism were about the carving out of Markov boundaries by the dynamics of free energy minimization (Friston, 2013). The variational framework, then, might allow us to model how organisms extend their Markov blankets into the environment, at a host of different spatial and temporal scales (Ramstead et al., 2019); e.g., to model the spider's web extending its ensemble of sensory states to include states outside its body (Kirchhoff & Froese, 2017).

In summary, the boundaries of cognitive systems are *nested* in that any system is made up of components, which (given that they, too, exist in a minimal sense) have a boundary that can be formalised as a Markov blanket. A given organism is essentially a hierarchical set of nested Markov blankets. Furthermore, there is a hierarchical listing of scales, in the sense that every state at a given scale is itself a mixture of blanket states at a smaller scale (see Figure 5 and Figure 6). The subsystems of interest here range from intra-cellular blanketed systems (e.g., organelles) to the blanket of the entire species. By the very fact that they are nested in this way, up to the scale of the species, the boundaries of any cognitive process of cognitive dynamics are *multiple*, in the sense that cognitive systems at different scales are integrated in one multiscale cognitive dynamics. The boundaries and scale that are *relevant* will depend on the kind of investigation we are aiming at, the phenomenon that is of interest, etc.

Concluding remarks: Towards multidisciplinary research heuristics for cognitive science

In this paper, we have attempted to overcome a common tendency to think of the boundaries of cognitive systems as either brain-based, embodied, or ecological/environmental by appealing to a multiscale interpretation of Markov blankets under the variational FEP. The

resulting *multiscale integrationist* perspective suggests that the boundaries of cognition are multiple and nested.

Some of the radical externalist views on cognition that we have discussed suggest that the divide between internalism and externalism is problematic (Thompson & Stapleton, 2009). We agree, precisely because each of these two options begs the question over where to look for the realisers of cognition. We argued in favour of an *ontological pluralism* based on a multiscale formulation of Markov blankets under the FEP. We argued that ontologically, states statistically insular or segregated at one scale are integrated by the dynamics (i.e., adaptive behaviour) at scales above. States separated by their respective Markov blankets are dynamically and statistically linked as states of the same higher-order system. The recursively nested, multilevel formulations of the Markov blanket formalism under the FEP allow to study the realisers of cognition, while acknowledging that they are a moving target; they shift according to the level of inquiry.

Some radical externalist views, enactive approaches especially, cast cognition as a relational phenomenon that equally recruits states of the brain, the body, and the world. The view we propose here agrees with the relational aspect of this project, but rejects the a priori emphasis in the assumption that all factors contribute equality to the causal patterns of interest. That cognition is inherently relational, that it integrates the contribution of states that are internal (systemic) and external to any given boundary, does not imply that the contributions of all relevant components are equal. Certain kinds of cognition rely mainly on the contributions of internal states (e.g., mental calculations); other activities are more embodied, and rely mainly on physiological or morphological states (e.g., walking); and other still depend most on the influence of abiotic, environmental factors or culturally patterned practices (e.g., driving a car). The approach we advocate here casts cognition as radically relational at each scale, even within the brain; e.g., relations between cells, relations to the brain's microenvironment, between different networks or again, between different patterns of functionally integrated units; without for all that endorsing the view that nothing matters more than anything else. This speaks to the necessity of *methodological pluralism* in cognitive science; and to the importance of developing new interdisciplinary research heuristics to determine and study, for any phenomenon, the relevant levels of description that are necessary to account for it.

Our multiscale integrationist formulation of the boundaries of cognition rejects any kind of essentialism about the boundaries of cognition. It suggests that explanations of cognition will differ conditioned on the phenomenon and our explanatory interests. In this

sense we are more aligned with Clark (2008) when he encourages us to “let a thousand flowers bloom” (p. 117). However, we restrict the scope of this gardening project by arguing that the FEP plays a coordinating and constraining role on the kind of explanations one should be looking for in the cognitive sciences.

References

Allen, M. (2018). The Foundation: Mechanism, Prediction, and Falsification in Bayesian Enactivism. Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Allen, M., & Friston, K. J. (2016). From cognitivism to autopoiesis: towards a computational framework for the embodied mind. *Synthese*, 1-24.

Anderson, M. (2017). Of Bayes and bullets: An embodied, situated, targeting-based account of predictive processing. *Philosophy and predictive processing. Frankfurt am Main: MIND Group*.

Badcock, P. B., Friston, K., & Ramstead, M. (2019). The Hierarchically Mechanistic Mind: A free-energy formulation of the human psyche. . *Physics of life Reviews*.

Bruineberg, J., Kiverstein, J., & Rietveld, E. (2016). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese*, 1-28. doi:doi:10.1007/s11229-016-1239-1

Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in human neuroscience, 8*. doi:doi.org/10.3389/fnhum.2014.00599

Bruineberg, J., Rietveld, E., Parr, T., van Maanen, L., & Friston, K. J. (2018). Free-energy minimization in joint agent-environment systems: A niche construction perspective. *Journal of Theoretical Biology, 455*, 161-178.

Campbell, J. O. (2016). Universal Darwinism as a process of Bayesian inference. *Frontiers in Systems Neuroscience, 10*.

Chemero, A. (2009). Radical embodied cognition. In: Cambridge, MA: MIT Press.

Christopoulos, G. I., & Tobler, P. N. (2016). Culture as a response to uncertainty: foundations of computational cultural. *The Oxford handbook of cultural neuroscience, 81*.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. New York: Oxford University Press

Clark, A. (2017). How to knit your own Markov blanket. In T. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing*. Frankfurt am Main: MIND Group.

Clark, A., & Chalmers, D. (1998). The extended mind. *analysis, 58*(1), 7-19.

Conant, R. C., & Ashby, W. R. (1970). Every Good Regulator of a system must be a model of that system. *Int. J. Systems Sci., 1*(2), 89-97.

Constant, A., Bervoets, J., Hens, K., & Van de Cruys, S. (2018). Precise worlds for certain minds: An ecological perspective on the relational self in autism. *Topoi*.

Constant, A., Ramstead, M., Veissière, S., Campbell, J., & Friston, K. (2018). A variational approach to niche construction. *Journal of the Royal Society Interface*.

Dawkins, R. (1982). *The Extended Phenotype: The Long Reach of the Gene*. Oxford: Oxford University Press.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences, 6*(4), 485-507.

Di Paolo, E. (2009). Extended life. *Topoi, 28*(1), 9.

Di Paolo, E., Buhrmann, T., & Barandiaran, X. (2017). *Sensorimotor life: An enactive proposal*: Oxford University Press.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences, 4*(4), 429-452.

Di Paolo, E. A., & Thompson, E. (2014). The enactive approach. In L. E. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 68-78). London: Routledge.

Dissanayake, E. (2009). The artification hypothesis and its relevance to cognitive science, evolutionary aesthetics, and neuroaesthetics. *Cognitive Semiotics, 5*(fall2009), 136-191.

Engel, A. K., Friston, K. J., & Kragic, D. (2016). The pragmatic turn: Toward action-oriented views in cognitive science.

Fabry, R. E. (2017). Betwixt and between: the enculturated predictive processing approach to cognition. *Synthese*, 1-36.

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in human neuroscience, 4*, 215.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127-138.

Friston, K. (2012). A free energy principle for biological systems. *Entropy, 14*(11), 2100-2121.

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

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2016a). Active inference and learning. *Neuroscience & Biobehavioral Reviews, 68*, 862-879.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active Inference: A Process Theory. *Neural Comput, 29*(1), 1-49. doi:10.1162/NECO_a_00912

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological cybernetics, 104*(1-2), 137-160.

Friston, K. J. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences, 360*(1456), 815-836.

Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2016b). Active inference: a process theory. *Neural computation, 29*, 1-49.

Friston, K. J., Levin, M., Sengupta, B., & Pezzulo, G. (2015). Knowing one's place: a free-energy approach to pattern regulation. *Journal of the Royal Society Interface, 12*(105).

Friston, K. J., Parr, T., & de Vries, B. (2017). The graphical brain: belief propagation and active inference. *Network Neuroscience, 1*(4), 381-414.

Friston, K. J., Rosch, R., Parr, T., Price, C., & Bowman, H. (2018). Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews, 90*, 486-501.

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese, 159*(3), 417-458.

Froese, T., & Di Paolo, E. A. (2011). The enactive approach: Theoretical sketches from cell to society. *Pragmatics & Cognition, 19*(1), 1-36.

Gallagher, S. (2006). *How the body shapes the mind*: Clarendon Press.

Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*: Oxford University Press.

Gallagher, S., & Allen, M. (2016). Active inference, enactivism and the hermeneutics of social cognition. *Synthese*, 1-22. doi:http://doi.org/10.1007/s11229-016-1269-8

Gibson, J. J. (1979). *The ecological approach to visual perception*: Psychology Press.

Gładziejewski, P. (2016). Predictive coding and representationalism. *Synthese, 193*(2), 559-582.

Gładziejewski, P., & Miłkowski, M. (2017). Structural representations: causally relevant and different from detectors. *Biology & Philosophy*, 1-19.

Hesp, C., Ramstead, M., Constant, A., Badcock, P. B., Kirchhoff, M., & Friston, K. (2019). A multi-scale view of the emergent complexity of life: A free-energy proposal. In M. P. e. al. (Ed.), *Evolution, Development, and Complexity: Multiscale Models in Complex Adaptive Systems*: Springer.

Hohwy, J. (2014). *The predictive mind*. Oxford: Oxford University Press.

Hohwy, J. (2016). The self-evidencing brain. *Noûs, 50*(2), 259-285.

Hohwy, J. (2017). How to Entrain Your Evil Demon. *Philosophy and predictive processing. Frankfurt am Main: MIND Group*.

Hohwy, J., Roepstorff, A., & Friston, K. J. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition, 108*(3), 687-701.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT press.

Hutchins, E. (2010). Cognitive ecology. *Topics in cognitive science, 2*(4), 705-715.

Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. Cambridge, MA: MIT Press.

Ingold, T. (2001). From the transmission of representations to the education of attention. *The debated mind: Evolutionary psychology versus ethnography*, 113-153.

Kiefer, A., & Hohwy, J. (2017). Content and misrepresentation in hierarchical generative models. *Synthese*, 1-29.

Kirchhoff, M. (2015). Species of realization and the free energy principle. *Australasian Journal of Philosophy, 93*(4), 706-723.

Kirchhoff, M. (2018a). Autopoiesis, free energy, and the life–mind continuity thesis. *Synthese, 195*(6), 2519-2540.

Kirchhoff, M. (2018b). Hierarchical Markov Blankets and Adaptive Active Inference : Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Kirchhoff, M. (2018c). Predictive processing, perceiving and imagining: Is to perceive to imagine, or something close to it? *Philosophical Studies, 175*(3), 751-767.

Kirchhoff, M., & 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., & Kiverstein, J. (2019). *Extended Consciousness and Predictive Processing: A Third-Wave View*. New York Routledge.

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface, 15*(138), 20170792.

Kirchhoff, M., & Robertson, I. (2018). Enactivism and predictive processing: a non-representational view. *Philosophical Explorations, 21*(2), 264-281.

Kirchhoff, M. D. (2012). Extended cognition and fixed properties: steps to a third-wave version of extended cognition. *Phenomenology and the Cognitive Sciences, 11*(2), 287-308.

Menary, R. (2010). The Extended Mind and Cognitive Integration. In R. Menary (Ed.), *The extended mind*. Cambridge, MA: MIT Press

Noë, A. (2004). *Action in perception*: MIT press.

Parr, T., & Friston, K. J. (2017). Uncertainty, epistemics and active inference. *Journal of the Royal Society Interface, 14*(136), 20170376.

Pearl, J. (1988). Probabilistic reasoning in intelligent systems: Networks of plausible inference. In. San Mateo, CA: Morgan Kaufmann.

Queller, D. C. (2014). Joint phenotypes, evolutionary conflict and the fundamental theorem of natural selection. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 369*(1642), 20130423.

Ramstead, M., Badcock, P. B., & Friston, K. (2018a). Answering Schrödinger's question: A free-energy formulation. *Physics of life Reviews, 24*, 1-16.

Ramstead, M., Badcock, P. B., & Friston, K. (2018b). Variational neuroethology: Answering further questions: Reply to comments on "Answering Schrödinger's question: A free-energy formulation". *Physics of life Reviews, 24*, 59-66.

Ramstead, M., Constant, A., Badcock, P. B., & Friston, K. (2019). Variational ecology and the physics of sentient systems. *Physics of life Reviews*.

Ramstead, M., Veissière, S., & Kirmayer, L. (2016). Cultural affordances: scaffolding local worlds through shared intentionality and regimes of attention. *Frontiers in Psychology, 7*.

Schwartenbeck, P., & Friston, K. (2016). Computational phenotyping in psychiatry: a worked example. *eneuro*, ENEURO. 0049-0016.2016.

Seth, A. K. (2014). The cybernetic brain: from interoceptive inference to sensorimotor contingencies. In T. Metzinger & J. M. Windt (Eds.), *Open MIND* (pp. 1-24). Frankfurt am Main: MIND Group.

Shipp, S. (2016). Neural Elements for Predictive Coding. *Front Psychol, 7*, 1792. doi:10.3389/fpsyg.2016.01792

Stotz, K. (2010). Human nature and cognitive–developmental niche construction. *Phenomenology and the Cognitive Sciences, 9*(4), 483-501.

Sutton, J. (2007). Batting, habit and memory: The embodied mind and the nature of skill. *Sport in Society, 10*(5), 763-786.

Sutton, J. (2010). Exograms and Interdisciplinarity: History, the Extended Mind, and the Civilizing Process. In R. Menary (Ed.), *The extended mind* (pp. 189-225). Cambridge, MA: MIT Press.

Thompson, E. (2010). *Mind in life: Biology, phenomenology, and the sciences of mind*: Harvard University Press.

Thompson, E., & Stapleton, M. (2009). Making sense of sense-making: Reflections on enactive and extended mind theories. *Topoi, 28*(1), 23-30.

Van de Cruys, S., & Wagemans, J. (2011). Putting reward in art: a tentative prediction error account of visual art. *i-Perception, 2*(9), 1035-1062.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*: MIT press.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological functions. In. Cambridge, MA: Harvard University Press.

Williams, D. (2017). Predictive processing and the representation wars. *Minds and Machines*, 1-32.

Yufik, Y. M., & Friston, K. (2016). Life and Understanding: the origins of "understanding" in self-organizing nervous systems. *Frontiers in Systems Neuroscience, 10*.

Zeki, S. (2005). The Ferrier Lecture 1995 behind the seen: The functional specialization of the brain in space and time. *Philos. Trans. R. Soc. Lond. B. Biol. Sci., 360*(1458), 1145–1183.

Zeki, S., & Shipp, S. (1988). The functional logic of cortical connections. *Nature, 335*, 311-317.

9. A comprehensive scholarly discussion of all the findings

9.1. Overview

Generally speaking, the main outcomes of the work presented in this doctoral thesis are twofold. First, the papers in this thesis collectively proposed a model of explanation and a formalism that are apt to enable and guide the study of multiscale systemic dynamics. This model of explanation and formalism rest on three integrative theoretical frameworks for studying cognitive systems at and across spatiotemporal scales. Second, this thesis proposed an argument, based on this model of explanation and formalism, against essentialism about the boundaries of cognitive systems in the philosophy of the cognitive sciences.

The three frameworks that were proposed in this thesis are: (1) a theoretical framework for understanding and modelling human sociocultural action, cognition, and learning, and the ways in which culture permeates possibilities for human action, called the *cultural affordances framework*; (2) a novel theoretical approach to neuroethology, which is the study of the mechanics and coevolution of adaptive behaviour and structures that control it (typically brains), called *variational neuroethology*; and (3) a new approach to cognitive ecology, that is, the study of how organisms and environments adapt to one another across time. This provides a theoretical framework to draw the boundaries around higher-order cognitive systems – and to formulate a physics of sentient systems – called *variational ecology*. Taken together, these three frameworks constitute an explanatory model that cognitive scientists can leverage to study the dynamics of systems that exist at and across the spatial and temporal scales at which they are manifest.

The main *philosophical* outcome of this thesis is an argument against essentialist conceptions of the boundaries of cognitive systems. That is, by leveraging the above model and formalism, one can overcome conceptions of cognitive systems that privilege one single type of boundary over others. The thesis advances ontological and methodological pluralism as a consequence. Ontological pluralism means that the boundaries of cognitive systems must be understood as multiple and nested, and that which among these boundaries is the most salient to cognitive scientists will be determined relative to their interests. Methodological pluralism is a corollary of the ontological kind; and means that, given the multitude and heterogeneity of cognitive boundaries, the best course of action to study cognitive systems will inevitably have to draw from different methodologies, from neuroimaging and genetic sequencing to behavioural measures and ethnographic techniques – and on various mathematical and statistical techniques to integrate and interpret these measures.

The outcomes of my doctoral research have bearings on theoretical work in the cognitive sciences and for the simulation of the dynamics of cognitive systems, and also pertain to scientific research culture. The cultural affordances framework, variational neuroethology, and variational ecology are (meta-) theories of cognitive systems that can provide an integrative, fully generalizable framework enabling the study and modelling of the dynamics and interactions of nested cognitive systems embedded in their ecological niche – from microscopic phenomena such as cellular and subcellular processes, to mesoscopic phenomena, such as neural dynamics, homeostasis, and adaptive action; to macroscopic phenomena, such as niche construction and evolution by natural selection.

In this final section, I will discuss the main implications of the work undertaken in this thesis, and unpack what I take to be the main limitations of this work. The first implication concerns the *formal ontology* that was developed in the last three chapters of the thesis, and how it relates to the philosophical debate between internalism and externalism about the boundaries of cognitive systems in the philosophy of the cognitive sciences, rather than engaging with the debate between emergentism and reductionism. The second concerns the limits and explanatory scope of the variational framework that was developed in the second and third chapters. The third concerns the limitations and blind spots of the approach to culture and cognition that was developed in the first chapter. The fourth set of considerations regards the relations between my doctoral project and phenomenological philosophy.

9.2. A formal ontology for multiscale systems: Against *a priori* metaphysics of emergence, and towards a naturalistic ontology of nested systems

The work presented in this thesis was premised on the notion that the debate between internalism and externalism in the philosophy of the cognitive sciences was more prescient – to clarify what was at stake in the study of multiscale systemic dynamics – than the debate between emergentism and reductionism. The choice to focus on this debate (instead of engaging the closely related debate over the epistemological status of downward causation and the metaphysics of emergent properties and their relation to the basal phenomena from which they emerge) may seem a bit contentious. This section will attempt to justify this choice and to consider the implications of the arguments and frameworks afforded by this thesis.

One of the main theoretical results of the work in this thesis is a *multiscale formal ontology of cognitive systems*. An ontology *simpliciter*, in this context, is a meta-theory that

tells us about the kinds of things that exist according to a given target theory – that is, a typology of the kinds of systems that exist and that can be studied using the resources of that theory. A *formal* ontology is the application of a formalism (typically, a logical, mathematical, statistical formalism) in order to answer the kinds of questions that are raised by ontology. Typically, in philosophy, the task of constructing an ontology, that is, of determining the kinds of things that exist, what we can say about them, and the criteria for 'thinghood' – as it were – are left to conceptual analysis. More precisely, these questions are typically addressed in an *a priori* fashion by metaphysics. To use a formalism to answer ontological questions is not unheard of. Philosopher nowadays turn to metaphysics to answer questions of this kind. Much work in metaphysics – classics as well as more contemporary work – concerns the application of logical formalisms, especially modal logic, to make sense of ontological issues (Lewis, 1986; Quine, 1961; Williamson, 2013). This work in analytic metaphysics has become a standard and popular way to deal with issues surrounding the development and articulation of particular ontologies. For example, much of the work done on emergent phenomena has been about specifying the right kind of logical relation between emergent (typically, mental) and basal (typically, physical) phenomena.

Much of this work has focused on varieties of the *supervenience* relation (Davidson, 1980/1970; Kim, 1993, 1999, 2006; Yoshimi, 2007, 2011, 2012). Briefly, supervenience is defined as a relation between properties or events, say A and B, where A supervenes on B if and only if there can be no changes in A without a corresponding change with respect to B. This might seem like a relation adequate to model relations between nested systems. However, upon reflection, while it may be adequate to capture relations between mental states (or events or properties) and physical ones, it almost invariably commits one to the causal inertness of supervening things on basal ones (Kim, 1989) – a consequence that we have tried to avoid in this thesis. It seems like supervenience is not specific enough to capture exactly the kind of *nesting* relation that we need to address multiscale causal dynamics. Supervenience is a purely *logical* construction, the relata of which can be very abstract properties, bearing little connection to the concerns of cognitive scientists – and it is not specifically tailored to address the kinds of issues that arise in the study multiscale systemic dynamics. Consider, for instance, the property of being *F*, whatever *F* may be. For any property *F*, the property of being not-*F* supervenes on that of being *F* – since, trivially, there cannot be a change with respect to *F*-ness without a change with respect not-*F*-ness. This hardly seems like a basis specific enough to link systemic phenomena across scales.

The formal ontology that is proposed in this thesis was first articulated in the second chapter, and was one of the two main components of the *variational neuroethology* that was proposed there – the other component being a heuristic for research on adaptive behaviour in nested multiscale systems. Two of the commentaries on our paper raised questions about how to draw the boundaries of Markov blanketed systems beyond the organism (Bruineberg & Hesp, 2018; Kirmayer, 2018). This challenge was taken up by the third chapter.

The multiscale formal ontology that we propose answers many of the questions that need to be addressed to study multiscale systemic dynamics. First, our ontology answers ontological question par excellence, that of *existence*: What is it to exist? – And what does it mean to count as a thing? The supervenience formalism presupposes an account of individualization, whereas our account starts from one. Via our formal ontology, this question is reinterpreted as the requirement that a system be endowed with conditional independence from its embedding environment – which, under the variational approach, is operationalized as the requirement that a system be endowed with a Markov blanket. This provides, via a statistical formalism, a rigorous and unambiguous answer to a traditionally metaphysical question. More importantly perhaps, although it draws on the Markov blanket formalism to formulate this answer, any *specific* answer to the existence question will inevitably be based on the facts of experience – i.e., be based on our observations and measurements regarding the conditional independence of some phenomenon of interest (or lack thereof). After all, what determines whether a system counts as a system under this formalism is the verifiable (indeed, observable, measurable) fact that it is endowed with conditional independence with respect to its embedding environment.

Under the FEP, if a system exists in a statistical sense, then it must have a Markov blanket – indeed, if it exists, then we can always define the appropriate blanket states. That is, the framework provided by variational neuroethology and variational ecology provides a principled answer to the question: What is it for a set of interacting things to be (to exist as) a system? Given this answer, one now disposes of a principled way to answering the other central metaphysical question: Which systems exist? So, the variational framework provides an ontology of systems (which is a good thing to have in the cognitive sciences – and in closely related sciences such as biology, anthropology, and the social sciences – since they are sciences of systems. Importantly, however, this is not a general metaphysics, since its does not allow us to answer questions about things that are not systems: time, events, art, numbers, etc. As a guide to research on cognitive systems, this seems more generative and

empirically adequate than do *a priori* pronouncements of the sort one might find in metaphysics.

Second, and most importantly for the study of multiscale cognitive systems, the formal ontology proposed in this thesis deals explicitly with the status of higher-order emergent phenomena and the question of how they are related to the patterns from which they emerge. Again, the approach on offer here eschews metaphysics, being derived entirely from the (observable, measurable) statistical structure of the phenomena being studied. The idea on offer is reminiscent of the philosopher Daniel Dennett's notion of 'real patterns' (Dennett, 1991). This is the idea that a given thing is real if and only if it corresponds to a real pattern in the world – that is, some patterned movement of matter that has real, measurable effects. The basic idea behind the *multiscale* formal ontology that is proposed in this thesis is simple, and continuous with this idea. It is, in a nutshell, that any component part of a Markov blanketed thing – insofar as it is *also* a thing that exists – must have a Markov blanket as well, by virtue of what it means to be a thing at all. This intuition fuels the extension of the formal ontology to cover *systems of systems* – and to make sense of dynamics all the way up and all the down the nested series of systems. Indeed, this intuition about scales lies at the heart of the approach proposed here. Markov blankets at a higher scale are statistical constructs, the states of which fluctuate more slowly (that is, at a slower timescale) than the internal states that they enshroud.

This multiscale formal ontology effectively deals with the issues pertaining to the ontological status of higher-order systems, paving the way to a systematic approach to their study – all the while avoiding the pitfalls of an *a priori* approach. On this reading, emergent phenomena exist if and only if they can be associated with a Markov blanket at a 'higher' spatial and temporal scales (that is, over slower temporal scales and vaster spatial scales). This formal, statistical approach to the questions raised by ontology has the advantage of being based on the (observable and measurable) statistical structure of the phenomena being studied, rather than on *a priori* metaphysical categories – again, consistent with Dennett's real patterns, the formalism tells us that if some higher-order pattern evinces a sufficient degree of conditional independence, then it counts as a system at higher spatial and temporal scales.

As discussed in the introduction, the *a priori* metaphysics of emergence is not a fruitful approach to study multiscale systemic dynamics – for two main reasons. First, there exist formal means to articulate emergentist (synergetic) and reductionist (renormalization group theoretic) accounts of the relation between higher-order phenomena and the

phenomena from which they emerge. It may be that different kinds of systems call for different ontological assumptions – a perspective easily accommodated by our own formal ontological approach, but not by approaches that privilege *a priori* resolution of these issues. Second, it seems plain to me that questions about the nature of emergent phenomena should not be decided *a priori* – and especially not via conceptual analysis, which is not sufficiently constrained by empirical fact. That is, we are not rejecting metaphysics and ontology, but rather the *a priori* forms of these projects; and favour a naturalistic approach instead. Rather, these questions should be arbitrated by the empirically observable and measurable facts of the matter; and any formalism to which we appeal should give us a better grip on these facts, instead of speaking in their stead. We are, after all, trying to talk about the dynamics of systems that exist independently of our conceptual schemes – our concepts might guide us in our investigation, but in the end, nature is quite indifferent to our conceptual categories.

As Pierre Poirier has emphasised in our discussions, this lends the theoretical framework developed in this thesis a rather 'meta' flavour, since it applies *to itself*. Metaphysics conducted in an *a priori* fashion (and indeed, perhaps also natural selection) have given us priors about what we think exists. A naturalistic epistemology based on variational neuroethology and variational ecology, however, affords something that *a priori* metaphysics does not; namely, a way of correcting these priors given evidence – that is, a Bayesian naturalistic metaphysics, i.e., something akin to empirical Bayes, but for metaphysics and ontology.

From the point of view of the theory that we advocate, metaphysical *a priori* categories are something like Bayesian priors. They may represent a good-enough guess as to what is going on. Indeed, as hermeneutic philosophy has long argued (Gadamer, 2010/1927; Heidegger, 2010/1927), we only ever approach the world from the vantage point or horizon provided by our priors beliefs, expectations, and ways of being in the world. However, it is important to tune our priors to the environment – and at times, to reject them if they fail, lest we fall into madness (Frith & Friston, 2013). As research on delusions and schizophrenia has abundantly made clear, if our priors are off target, they may do more damage than good. I believe that the tradition of emergentist philosophy has led us a *levels-based* ontology, an ontology of distinct forms of existence (e.g., atomic existence, cellular existence, organismic existence, all related by logical relations) – where a *process-based* ontology, like the one proposed, would be more fruitful. Indeed, active inference is a story about the emergence and maintenance of Markov blankets – the ontology in play follows directly from the *process of existing*. We might speculate that much of the conceptual machinery used to study such

phenomena inherit from the tradition of British emergentist ontologies (Ablowitz, 1939; McLaughlin, 1992). These traditional views posit a *hierarchy of levels*, from atoms and molecules, to cells and organs, to organisms and social systems – each level existing as relatively independent. Strong emergentist views are certainly continuous with this view. Focusing on conditional independence and the ways that it is established and maintained through active inference is, I take it, a great source of *prediction errors* in this respect – which will help making our study empirically tractable and appropriately reactive to mistakes.

Third, the approach here frays a path *between* weak and strong readings of emergence. It is close in spirit to weak (epistemological) readings of emergence because it remains agnostic about the *metaphysics* of downward causation. Indeed, on the formal ontology proposed here, variational free-energy is generated at each scale by the relevant patterns, and is integrated across scale in the selection of action policies. We thus have an account of how emergent phenomena can be causally efficacious – in effect, their role is to establish probability gradients on the trajectories of the constituent parts. Importantly, this does not commit us to a view on the *metaphysics* of *inter-level* causation; not least because we are replacing the *metaphysics of levels of being* with an approach to phenomena that are integrated systems of systems spanning several *spatial and temporal scales*. In summary, if successful, the formal ontology proposed by this thesis effectively supersedes metaphysical approaches to multiscale dynamics.

9.3. The limits of the variational approach to the cognitive sciences

Since the variational approach appears so all encompassing – or very nearly so – it may be worthwhile to discuss briefly its explanatory scope and to point out a few of its limits. First of all, as is emphasised throughout the thesis, the free-energy principle is *not* a theory of everything. Rather, it is a theory of *biological and cognitive self-organisation via adaptive action*. More technically, it is a fully general account of the constraints that must be met by any system that exists at nonequilibrium steady state. That is, it describes the most general conditions that must be fulfilled by any system with a set of attracting states, or a phenotype.

The explanatory scope of the FEP is clearly very wide, but this breadth is purchased at the cost of saying, in effect, very little about the specifics of the biological systems so described or explained. This is not a defect of the principle – at least, arguably not. Indeed, as we argue in the last three chapters of this thesis, such minimalism positions the free-energy principle to act not as a finished theory of everything, but instead to provide a general

theoretical and mathematical framework to enable and guide multidisciplinary research on the behaviour of adaptive systems – especially in the multiscale formulations thereof. As Axel Constant has argued in discussions, the specificity of FEP-based explanations of given phenotypes rests on the specificity of the *generative models* that describe those phenotypes – that is, the set of beliefs about the typical causes of sensory states that make a creature 'the kind of creature that it is' (Friston & Buzsáki, 2016; Friston, Redish, & Gordon, 2017; Parr & Friston, 2018). One might say, then, that the species-specific character of a given FEP-based explanation can this coexist happily with the generality of the FEP as an explanatory principle.

The second and third chapters of this thesis makes this very clear, in principle and in practice. In principle, the variational neuroethology and variational ecology that were developed there essentially and irreducibly draw upon established frameworks in biology, namely evolutionary systems theory and Tinbergen's four questions (mechanism, ontogeny, phylogeny, and function) to propose a general account of the adaptive behaviour of multiscale systems and their control systems. Here, the FEP augments and complements existing bodies of theory, allowing for theoretical integration. In practice, we also derived a specific theory of the embodied, situated human brain from the more meta-theoretical work of variational neuroethology and variational ecology, namely, the hierarchically mechanistic mind – which we have continued to develop since (Badcock, Friston, & Ramstead, 2019). This theory appeals to theories that specifically address the human brain in its idiosyncrasy; it appeals to human neuroscience as well as other disciplines like human cognitive psychology and cognitive-evolutionary anthropology.

To talk about a given species meaningfully and scientifically, it is necessary to look into the *specific historical trajectories* of the systems we are studying. This is why we appealed to specific, substantial explanations in the second chapter in constructing our approach to neuroethology. In the end, *historicity* may be the defining limitation of approaches to cognitive systems based on the FEP. The FEP is formulated in terms of *conservative* dynamics. That is, it captures the dynamical constraints that must necessarily be in play in systems that are endowed with a phenotype. The FEP tells us what must be true a system with a phenotype for it to remain in (that is, to *periodically return to*) the states that make it 'the kind of creature that it is'. It is far from clear that *historical drift and change* can be accounted for.

These are not ultimate limitations of the FEP, they have more to do with the mathematical assumptions that are in play in the current versions of the construction. One

such assumptions is worth mentioning here: The FEP holds for systems that are *locally ergodic* – which means, very crudely, that the average of our measures of that system coincide, over a long enough time, with a measure of the average state of such a system. While this is a plausible assumption for most biological systems, it may not be adequate for systems the dynamics of which are extremely transient; i.e., it may be a fair assumption to model societies in times of stability (e.g., feudal France, circa 1500) but not in times of great change (e.g., the French revolution and ensuing terror).

Intriguingly, the FEP is deeply related to (indeed, can be derived from) what may be a *theory of every* '*thing*' – in the sense of being a theory that can be applied to anything that is a thing at all, i.e., a physics of anything that is endowed with robust conditional independence from its embedding environment. The Markov blanket formalism, it has recently been shown, can be extended to cover all sorts of phenomena, ranging quantum to classical mechanics and thermodynamics (Friston, in preparation). Although the details of this ambitious proposal are far beyond the scope of this thesis, the existence of such a project does makes the FEP a *special case* of a more general theory of every 'thing' – as applied to systems the dynamics of which are dependent on their internal and active states.

9.4. The cultural affordances framework: Affording blind spots

The cultural affordances framework contributes to our understanding of multiscale dynamics by formulating a mechanism of the enculturation of human agents that involves factors pitched at different descriptive levels. We followed the lead of cognition, communication, and culture researcher Andreas Roepstorff, who suggested that the correct unit of analysis for studying cultural phenomena was that of patterned cultural practices (Roepstorff, Niewöhner, & Beck, 2010). Patterned cultural practices were suggested as the ideal unit of analysis for studying the interaction between cognitive and cultural phenomena because is they represent a 'middle-range' level of analysis – ideally pitched in between the (overly coarse) macro-level of cultural groups and the (overly narrow) micro-level of encultured brain activity in individuals.

The first chapter identified one specific kind of patterned cultural practice that was central to enculturation – those practices that modulate the attention of human agents, which we call *regimes of attention*. The concept of regimes of attention is the central theoretical and mechanistic contribution of the first chapter, which ties together the various bodies of literature mobilized in the chapter. Indeed, the concept was used in the initial stages on my research as a case study on multiscale theorizing and modelling.

At the time of writing the first chapter, I had not yet become privy the integrative potential of the FEP to provide a general framework to explain multiscale integration. The construct of regimes of attention, when framed in terms of active inference, provide a natural and potent mechanism for the enculturation of human agents. The crucial outcome of this model, which makes it my first attempt to engage in multiscale theoretical pursuits, is that it combines factors operative at different spatial and temporal scales. According to the cultural affordances framework, regimes of attention organise and recruit practices and material artefacts at one scale – i.e., the scale of the embedding social and cultural environment. The effects of regimes of attention is to pattern the enculturation (that is, the culture-guided development) of human agents at another scale (at the scale of ontogeny), such that immersive participation in regimes of attention ends up installing in the learner (roughly) the same expectations that guided the practices in which they were immersed to begin with. This process itself is the result of accumulated interactions with the environment, which unfold over real-time or mechanistic timescales. Thus, true to its original aim, the cultural affordances framework provides a model of human sociocultural action and cognition that integrates factors from all these different scales.

The centrality of regimes of attention came to the fore through my collaboration with Samuel Veissière. The significance of these kinds of cultural practices was made especially salient since I was also working on active inference models of *attention* while writing this chapter. Under active inference, the construct of attention has mostly been cast as equivalent to the process of *precision weighting* – that is, the modulation of the gain or 'volume' of neural signals as a function of their estimated reliability or *precision* (Feldman & Friston, 2010). This process of precision weighting, in turn, is the main factor that determines what is retained by the system, since it effectively regulates belief updating in the brain. This feedback loop, when placed in cultural context, seemed like an ideal mechanism to explain enculturation. In other words, attention as precision weighting determines what content is processed – and that, in turn, determines what signals are processes and thereby learned by the agent.

The resulting framework was a good point of departure, but proved insufficient on its own. For the purposes of my doctoral research, the most significant limitation of the cultural affordances framework lies in its treatment of multiscale integration; namely, the framework lacked a model of explanation able to account for how the various causal factors to which we appealed at different levels of description are related and interact. On its own, the cultural affordances framework begs the question of how multiscale dynamics are realised and leaves

unanswered the question of how phenomena occurring at different levels of description are *integrated* into coherent patterns of adaptive behaviour. The following years of my doctoral work (from 2016 to 2019) were thus devoted to the development of a meta-theory that could justify the results of the first chapter. The integrative work of the subsequent chapters is premised on this search, and resulted in an attempt to systematically extend the FEP from a unified theory of the function, structure, and dynamics of the brain to a theory of the adaptive, ecologically situated behaviour of nested cognitive systems.

Another limitation of the cultural affordances framework, which was not apparent at the time of writing the paper but has become increasingly central to my more recent work, is that it does not capture one of the two intuitive or folk psychological meanings of the word 'attention'. Attention can refer to the metaphorical turning of the mind's eye to information that has *already* been received, and engaging with the information (via cognitive processing) as a function of their estimated relevance. As discussed, under the FEP, this is cast as the mechanism of *precision weighting*. But attention can also refer to 'eye-grabbing' nature of a phenomenon, as when we hear a loud sound and direct our gaze to its perceived source. This aspect of attention did not receive explicit treatment in the cultural affordances framework, but would play a critical role in the work that followed shortly thereafter. Beyond the phenomenological differences between the two kinds of attention, there is an intuitively obvious difference between choosing where to direct visual saccade to disambiguate a visual scene, on the one hand, and determining what to do with that information once it has been received.

This difference between these two forms of attention has been explored in the recent literature on active inference – as the difference between the process of precision weighting and the *epistemic affordance or salience* of an action policy (Parr & Friston, 2017). This latter quantity is involved in the selection of action policies, and quantifies the amount of uncertainty that is resolved by performing that action – as opposed to the utility that accrues to the performance of that action. Recent work on ecological interpretations of active inference cast epistemic affordance as the quantity that is optimised in the process of niche construction (Bruineberg, Rietveld, Parr, van Maanen, & Friston, 2018; Constant et al., 2018). On this reading, organisms are able to act on the basis of the shared value of an action policy by leveraging the traces left by the adaptive actions of conspecifics. The variational approach to niche construction that results from this perspective was one of the central components in the theoretical integration accomplished by the variational ecology that is proposed in the third chapter of this thesis. We have conducted follow-up work that further

develops this perspective as well (Constant, Ramstead, Veissière, & Friston, 2019; Veissière, Constant, Ramstead, Friston, & Kirmayer, Accepted).

9.5. Phenomenology: Reckoning with consciousness

The truth (if I'm being honest) is that the secret – and indeed, abandoned – target of this thesis was *phenomenology* (the study of *consciousness and conscious experience*) and hermeneutic philosophy (the study of the experience, process, and methods of interpretation).

One of the greatest lacunae of this doctoral thesis – and to me, its greatest failure – is the lack of a clear engagement with phenomenological philosophy, existentialism, and hermeneutics. The significance of this failure is so great in my view in no small part because the original intentions that motivated this doctoral work were deeply connected to phenomenological philosophy. My original hope in developing a multiscale approach to cognitive systems was to rehabilitate the descriptive level that corresponds to lived phenomenological experience.

I was originally inspired by the project for a cultural neuro-phenomenology that was put forward by my mentor at McGill University, Laurence Kirmayer. I had initially thought of my project as prolonging the work on the naturalization of phenomenology that I had conducted over the course of my Master's degree at Université du Québec à Montréal under the supervision of Pierre Poirier. Over the course of my Master's research, I had examined various forms of phenomenological philosophy and became deeply dissatisfied with classical, Husserlian styles of phenomenology. Husserlian theory was disappointing to me, and became impossible to work with, mainly because, despite the insistence of some to the contrary (Zahavi, 2003), Husserl's mature philosophy is a form of transcendental idealism – perhaps even a kind of transcendental absolutism, which reduced *everything* to the workings of transcendental subjectivity.

Coming from cognitive science, I was more influenced by the guiding idea of philosopher Paul Ricoeur – that against the phenomenological impetus to reduce everything *to* consciousness, we should instead pursue an alternative, hermeneutic aim: the reduction *of* consciousness; i.e., the situation of consciousness within the *deep history of processes* – causal and historical – that produced it. The resources of Heideggerian and Merleau-Pontian phenomenology were richer for my purposes. Heidegger's insistence on immediate engagement with the environment (Heidegger, 2010/1927) resonated deeply with the enactive approach that is employed throughout this thesis. The emphasis placed by Merleau-Ponty on embodied engagement with the world, and the emergence of meaning in embodied

engagement (Merleau-Ponty, 2013/1945), was perhaps the central guiding philosophical intuition of this thesis.

The work of Heidegger and Merleau-Ponty had already been leveraged to great effect in the cognitive sciences when I began my doctoral studies (Kiverstein & Wheeler, 2012; Thompson, 2010; Varela et al., 1991). The affordances construct – as revised by philosophers and cognitive scientists Tony Chemero, Jelle Bruineberg, Erik Rietveld, and Julian Kiverstein (Bruineberg et al., 2016; Bruineberg & Rietveld, 2014; Bruineberg et al., 2018; Chemero, 2003, 2009; Rietveld & Kiverstein, 2014) – was a central influence on my doctoral research. I followed their lead, using the resources of ecological psychology in an attempt to bring phenomenology into play, but in a way that was tractable to empirical inquiry (something that can hardly be said about Heidegger and Merleau-Ponty themselves).

Previous work by the aforementioned philosophers had provided a proof of concept that the affordances construct could effectively translate much of the Heideggerian insights about readiness-to-hand (*Zuhandenheit*) – that is, the mode of immediate engagement with things around us – within a framework that is empirically tractable and is able to guide inquiry in the cognitive sciences. From there, it made a great deal of sense to see whether it was possible to articulate a culturally richer account of affordances, informed by recent cognitive anthropology. This led to the development of the *cultural affordances* construct – based in no small part on notes and seminars by Samuel Veissière, who was working on a similar set of issues, but coming at things from anthropology, and drawing on thinkers such as Tim Ingold (e.g., Ingold, 2001) and aspects of Francisco Varela's work with which I had not previously engaged (Varela, 1999).

To my deep chagrin and discontent, the place of consciousness remains mysterious and unclear in the theoretical edifice that has resulted from my doctoral work. The multiscale framework sketched here, with its consideration of affordances, at least dialogues with phenomenology – minimally, to be sure. However, this is also where it fails: besides its being inspired, albeit very loosely, by phenomenological philosophy, there is no explicit engagement of conscious, first-person experience to be found in this work.

The meaning of the title of this thesis changed as I wrote my thesis. I originally wanted to say something about the *mind*, in the sense that occurs most in philosophy – i.e., the what-it-is-like or first-person, 'for-me' character of conscious experience. It was my contention that we had effectively lost sight of the phenomenological aspect of the mind in the cognitive sciences. Ironically, I also lost sight of consciousness as I developed the material in this thesis. The new meaning of title refers to the new question that became

central to the thesis: *Where* are we to draw the relevant boundaries of cognitive systems. In a sense, my thesis was transformed, from one question (Have we lost our *minds*?) to another (Have we *lost* our minds?).

My PhD supervisor Laurence Kirmayer has on numerous occasions pressed me to clarify the relation between the technical use of 'surprise' in the variational formalism with the folk psychological or phenomenological use of the term 'surprise', e.g., as referring to the subjective experience of being surprised. Surprise in information theory is a measure of the improbability of some state or outcome. In its folk psychological usage, 'surprise' emerges out of a wider and deeper set of informational processes with a particular (cultural) history, and implies an extensive social and temporal embedding related to consciousness, felt phenomenal qualities, and cultural meaning. For instance, psychologically, surprise occurs when some event happens that contradicts the psychological expectations had by an agent. To my disappointment (and doubtless, to my supervisor's as well), I have not been able to clarify the relation between the two uses of the term. Now, there must be some link between the two constructs. After all, at least heuristically, whenever some turn of events is psychologically surprising, this must be because there is some departure from expectations – and hence, some surprise. However, this seems like a sleight of hand and is not informative. After all, the quantity surprise is tracked by all cognitive processes, since variational free-energy is an upper bound on surprise, and all cognitive processes track (and avoid) variational free-energy. There is, then, a sense in which surprise is involved even when expectations are *met* at the psychological level. The two notions thus do not converge, and it is beyond my ability at this stage to determine how the two connect to one another.

My MA supervisor Pierre Poirier has suggested in discussions – more controversially, to be sure – that the frameworks that have been described in this thesis may serve to advance the agenda of *eliminative materialism* (Churchland, 1981). In a nutshell, eliminative materialism is the claim that our folk psychological concepts (such as the distinction between the mind and the body) are theories – and that like any theory, they are subject to revision. On this view, it also happens that folk psychology is a bad theory, which do not reflect the way that nature is. On this view, concepts such as 'mind' are akin to the ancient concept of 'elemental fire': they cover disparate phenomena bearing little relation to each other – the way 'fire' covered phenomena as unrelated as nuclear fusion (in the sun), bioluminescence (in insects and sea creatures), and rapid oxidation (in combustion).

The suggestion, then, is that the cultural affordances framework – when further developed to address a broader range of interactions between human agents and their

ecological niche – holds the promise of a wholesale *replacement* of folk psychological concepts (and indeed, perhaps also of many of the natural kind concepts from the ontologies of psychological and neuroscientific sciences, which inherit from folk psychology). This conceptual edifice provided by the cultural affordances framework arguably improves on folk psychology, in that the concepts in play always carve nature in a relational manner (in terms of organism-environment relations).

I find this suggestion intriguing – and, in some ways, comforting. The implication of this view, after all, is that the failings just discussed are not due to my own limitations – but instead follow from the fact that *there is no such thing as consciousness* as phenomenological philosophy understands it. Instead, there is something a diversity of experiences, each with their own phenomenology (in attention, in expectation, in narrative, in subjectivity, in feelings of agency, of mineness, of knowing, of existence, etc.). Perhaps some of these can be dealt with individually within the framework proposed in this thesis.

In the end, I hope my work in this thesis will lead others, who have more time and energy than I, to pursue this dialogue with the sciences that study first-person experience. I can only sketch general directions for future research here. My feeling is that there is much to accomplish within the broad, recently proposed project of a *neural hermeneutics* (Friston & Frith, 2015a; Friston & Frith, 2015b; Gallagher & Allen, 2016). This project is a revamped version of the attempt to bring phenomenology into contact with the natural sciences – what has been called ‘neuro-phenomenology’ (Gallagher, 2012; Ramstead, 2015; Zahavi, 2013). My only reservation with the neural hermeneutics project, which may not come as a surprise given the content of this doctoral thesis, is its overemphasis on neural dynamics. A more natural project, from the point of view articulated here, would be the project of *multiscale* hermeneutics, based on a multiscale reading of the FEP.

10. Final conclusion and summary

The aim of this doctoral thesis was to examine the question: How best to address the study of systemic dynamics that span several spatial and temporal scales in the cognitive sciences, from cells to societies? The thesis aims to prevent us from 'losing our minds', that is, losing sight of the full phenomenon of cognition by focusing unduly on phenomena at specific scales and levels of description to the detriment of others.

This thesis articulated a first principle approach to multidisciplinary research in the cognitive sciences. This synthesis is based on multiscale extensions of the variational free-energy principle, a framework from theoretical biology that provides an account of the general information-theoretic constraints that must be met by all cognizing organisms – and indeed, by any system with a phenotype. This provides the cognitive sciences with models of the dynamics of cognitive systems at, and across, all the scales at which they exist.

More specifically, leveraging work in philosophy and theoretical work in the cognitive sciences, this thesis proposed three new theoretical frameworks designed to address the dynamics of cognitive systems at and across scales: (1) the *cultural affordances framework*, an integrative theoretical model of how human action, cognition, learning, and culture, formulated in terms of possibilities for action in the environment (affordances) and the patterned cultural practices that direct the attention of human agents, leading to their enculturation (regimes of attention); (2) *variational neuroethology*, a new approach to the study of the mechanics and evolution of adaptive behaviour and behavioural control systems (such as brains), which foregrounds the construct of recursively nested cognitive systems (of systems); and finally, (3) *variational ecology*, a novel perspective on the project of cognitive ecology that emphasises the way such nested cognitive systems interact with the ecological niche in which they are embedded and with which they interact.

Finally, leveraging these new frameworks, the thesis developed a philosophical argument against what we labelled 'essentialism about the boundaries of cognitive systems', a debate which still rages in the philosophy of the cognitive sciences. Essentialist approaches are those that privilege one type of boundary over others, and which argue that the boundaries of cognitive systems are essentially of a specific kind. The thesis argued that the boundaries of cognitive systems are not fixed once and for all, and identical for all inquiries, but instead, that they are *multiple*, *nested*, and *interest-relative*. With this thesis, I hope to have helped to resolve the artificial and false dichotomy between internalism and externalism in the philosophy of the cognitive sciences. Having done so, the author hopes that the thesis will prevent us from losing our minds – *wherever* they happen to be.

11. Bibliography

Ablowitz, R. (1939). The theory of emergence. *Philosophy of Science, 6*(1), 1-16.

Aboud, F. E., & Amato, M. (2008). Developmental and socialization influences on intergroup bias. In S. M. Quintana & C. McKown (Eds.), *Handbook of race, racism, and the developing child* (pp. 65-85). Hoboken, NJ: John Wiley & Sons.

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology, 14*(1), 43-64.

Adams, F., & Aizawa, K. (2009). Why the mind is still in the head. *The Cambridge handbook of situated cognition*, 78-95.

Adams, F., & Aizawa, K. (2010). Defending the bounds of cognition. *The extended mind*, 67-80.

Adams, R. A., Bauer, M., Pinotsis, D., & Friston, K. J. (2016). Dynamic causal modelling of eye movements during pursuit: Confirming precision-encoding in V1 using MEG. *Neuroimage, 132*, 175-189.

Adams, R. A., Huys, Q. J. M., & Roiser, J. P. (2016). Computational Psychiatry: towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery, and Psychiatry, 87*(1), 53-63.

Allen, M. (2018). The Foundation: Mechanism, Prediction, and Falsification in Bayesian Enactivism. Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Allen, M., & Friston, K. J. (2016). From cognitivism to autopoiesis: towards a computational framework for the embodied mind. *Synthese*, 1-24.

Anderson, M. (2017). Of Bayes and bullets: An embodied, situated, targeting-based account of predictive processing. *Philosophy and predictive processing. Frankfurt am Main: MIND Group*.

Anderson, M. L. (2014). *After phrenology : neural reuse and the interactive brain*. Cambridge, MA; London, England: The MIT Press.

Anderson, M. L., & Chemero, T. (2013). The problem with brain GUTs: conflation of different senses of "prediction" threatens metaphysical disaster. *Behavioral and Brain Sciences, 36*(3), 204-205.

Anderson, M. L., & Finlay, B. L. (2014). Allocating structure to function: the strong links between neuroplasticity and natural selection. *Frontiers in human neuroscience, 7*.

Ao, P. (2003). Stochastic force defined evolution in dynamical systems. *arXiv preprint physics/0302081*.

Ao, P. (2005). Laws in Darwinian evolutionary theory. *Physics of life Reviews, 2*(2), 117-156.

Ao, P. (2008). Emerging of stochastic dynamical equalities and steady state thermodynamics from Darwinian dynamics. *Communications in theoretical physics, 49*(5).

Ao, P. (2009). Global view of bionetwork dynamics: adaptive landscape. *Journal of Genetics and Genomics, 36*(2), 63-73.

Ao, P., Chen, T.-Q., & Shi, J.-H. (2013). Dynamical decomposition of markov processes without detailed balance. *Chinese Physics Letters, 30*(7), 070201.

Apps, M. A. J., & Tsakiris, M. (2014). The free-energy self: a predictive coding account of self-recognition. *Neuroscience and Biobehavioral Reviews, 41*, 85-97.

Arnold, L. (2003). *Random Dynamical Systems (Springer Monographs in Mathematics)*. Berlin: Springer-Verlag.

Badcock, P. B. (2012). Evolutionary systems theory: a unifying meta-theory of psychological science. *Review of General Psychology, 16*(1), 10-23.

Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017a). The depressed brain: an evolutionary systems theory. *Trends in Cognitive Sciences, 21*, 182-194.

Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017b). The depressed brain: An evolutionary systems theory. *Trends in Cognitive Sciences, 21*(3), 182-194.

Badcock, P. B., Friston, K., & Ramstead, M. (2019). The Hierarchically Mechanistic Mind: A free-energy formulation of the human psyche. . *Physics of life Reviews*.

Badcock, P. B., Friston, K. J., Ramstead, M., Ploeger, A., & Hohwy, J. (accepted). The Hierarchically Mechanistic Mind: An evolutionary systems theory of the brain, mind and behavior. *Cognitive, Affective, and Behavioral Neuroscience*.

Badcock, P. B., Ploeger, A., & Allen, N. B. (2016). After phrenology: time for a paradigm shift in cognitive science. *Behavioral and Brain Sciences, 39*. doi:doi.org/10.1017/S0140525X15001557

Bak, P., Tang, C., & Wiesenfeld, K. (1988). Self-organized criticality. *Physical review A, 38*(1), 364-374.

Banaji, M. R., & Gelman, S. A. (2013). *Navigating the social world: What infants, children, and other species can teach us*. Oxford: Oxford University Press.

Bar, M. (2011). *Predictions in the brain: Using our past to generate a future*. Oxford: Oxford University Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577-660.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol., 59*, 617-645.

Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or surprise? *Frontiers in Psychology, 4*(907).

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron, 76*(4), 695-711.

Bechtel, W. (2007). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*: Psychology Press.

Bechtel, W. (2009). Looking down, around, and up: Mechanistic explanation in psychology. *Philosophical Psychology, 22*(5), 543-564.

Bechtel, W., & Abrahamsen, A. (2010). Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science Part A, 41*(3), 321-333.

Bedau, M. (1997). Weak emergence,. Mind, causation and world. *Philosophical Perspectives, vol-11*, 375-399.

Bedau, M. (2002). Downward causation and the autonomy of weak emergence. *Principia: an international journal of epistemology, 6*(1), 5-50.

Bedau, M. A., & Humphreys, P. E. (2008). *Emergence: Contemporary readings in philosophy and science*: MIT press.

Beer, R. D. (1995a). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence, 72*(1), 173-215.

Beer, R. D. (1995b). A dynamical systems perspective on agent-environment interaction. *Artificial intelligence, 72*(1-2), 173-215.

Bellomo, N., & Elaiw, A. (2018). Dynamics and equilibria of living systems: Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Dèsormeau Ramstead et al. *Physics of life Reviews*.

Bishop, R. C. (2008). Downward causation in fluid convection. *Synthese, 160*(2), 229-248.

Bitbol, M., & Gallagher, S. (2018). The free energy principle and autopoiesis: Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Blackmore, S. (2000). *The Meme Machine*. Oxford: Oxford University Press.

Boogerd, F. C., Bruggeman, F. J., Richardson, R. C., Stephan, A., & Westerhoff, H. V. (2005). Emergence and its place in nature: A case study of biochemical networks. *Synthese, 145*(1), 131-164.

Bourdieu, P. (1977). *Equisse d'une théorie de la pratique* Cambridge University Press.

Bourdieu, P. (1984). *Distinction: A Social Critique of the Judgement of Taste*: Harvard University Press.

Bracken, P., Thomas, P., Timimi, S., Asen, E., Behr, G., Beuster, C., . . . Double, D. (2013). Psychiatry beyond the current paradigm. *Psicoterapia e Scienze Umane, 47*(1), 9-22.

Broad, C. D. (2014/1925). *The mind and its place in nature*: Routledge.

Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence, 47*(1-3), 139-159.

Bruineberg, J. (2017). Active inference and the primacy of the 'I can'. In T. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing*. Frankfurt am Main: MIND Group.

Bruineberg, J., & Hesp, C. (2018). Beyond blanket terms: Challenges for the explanatory value of variational (neuro-) ethology: Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Bruineberg, J., Kiverstein, J., & Rietveld, E. (2016). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese*, 1-28. doi:doi:10.1007/s11229-016-1239-1

Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in human neuroscience, 8*. doi:doi.org/10.3389/fnhum.2014.00599

Bruineberg, J., Rietveld, E., Parr, T., van Maanen, L., & Friston, K. J. (2018). Free-energy minimization in joint agent-environment systems: A niche construction perspective. *Journal of Theoretical Biology, 455*, 161-178.

Buckley, C. L., Kim, C. S., McGregor, S., & Seth, A. K. (2017). The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology, 81*, 55-79.

Buckner, R. L., & Krienen, F. M. (2013). The evolution of distributed association networks in the human brain. *Trends in Cognitive Sciences, 17*(12), 648-665.

Byrne, R. W., & Whiten, A. (Eds.). (1988). *Machiavellian Intelligence: Social Expertise and Evolution of Intellect in Monkeys, Apes and Humans*. Oxford: Oxford University Press.

Campbell, D. T. (1974). 'Downward causation'in hierarchically organised biological systems. In *Studies in the Philosophy of Biology* (pp. 179-186): Springer.

Campbell, J. O. (2016). Universal Darwinism as a process of Bayesian inference. *Frontiers in Systems Neuroscience, 10*.

Campbell, J. O. (2018). Towards a unification of evolutionary dynamics: Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Caporael, L. R. (2001). Evolutionary psychology: Toward a unifying theory and a hybrid science. *Annual review of psychology, 52*(1), 607-628.

Carr, J. (1981). *Applications of Centre Manifold Theory*. Berlin: Springer-Verlag.

Changeux, J.-P. (2017). Climbing brain levels of organisation from genes to consciousness. *Trends in Cognitive Sciences, 21*(3), 168-181.

Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology, 15*(2), 181-195.

Chemero, A. (2009). Radical embodied cognition. In: Cambridge, MA: MIT Press.

Chiel, H. J., & Beer, R. D. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in neurosciences, 20*(12), 553-557.

Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. *Synthese, 191*(2), 127-153.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT press.

Choudhury, S., & Slaby, J. (2016). *Critical neuroscience: A handbook of the social and cultural contexts of neuroscience*: John Wiley & Sons.

Chow, J. Y., Davids, K., Hristovski, R., Araújo, D., & Passos, P. (2011). Nonlinear pedagogy: Learning design for self-organizing neurobiological systems. *New Ideas in Psychology, 29*(2), 189-200.

Christopoulos, G. I., & Tobler, P. N. (2016). Culture as a response to uncertainty: foundations of computational cultural. *The Oxford handbook of cultural neuroscience, 81*.

Churchland, P. M. (1981). Eliminative materialism and propositional attitudes. *The Journal of philosophy, 78*(2), 67-90.

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical transactions of the Royal Society B: Biological sciences, 362*(1485), 1585-1599.

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual review of neuroscience, 33*, 269-298.

Claidière, N., Scott-Phillips, T. C., & Sperber, D. (2014). How Darwinian is cultural evolution? *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 369*(1642), 20130368.

Clark, A. (1998). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT press.

Clark, A. (2004). *Natural-born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford: Oxford University Press.

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences, 10*(8), 370-374.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. New York: Oxford University Press

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181-204.

Clark, A. (2015a). Predicting Peace: The End of the Representation Wars-A Reply to Michael Madary. In T. W. Metzinger, J M (Ed.), *Open MIND*. Frankfurt am Main: MIND Group.

Clark, A. (2015b). *Surfing uncertainty: prediction, action, and the embodied mind*. Oxford: Oxford University Press.

Clark, A. (2017). How to knit your own Markov blanket. In T. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing*. Frankfurt am Main: MIND Group.

Clark, A., & Chalmers, D. (1998). The extended mind. *analysis, 58*(1), 7-19.

Clark, K. B. (1963). *Prejudice and your child*. Boston: Beacon Press.

Clark, K. B., & Clark, M. K. (1939). The development of consciousness of self and the emergence of racial identification in Negro preschool children. *The Journal of Social Psychology, 10*(4), 591-599.

Colombo, M. (2014). Explaining social norm compliance. A plea for neural representations. *Phenomenology and the Cognitive Sciences, 13*(2), 217-238.

Conant, R. C., & Ashby, W. R. (1970). Every Good Regulator of a system must be a model of that system. *Int. J. Systems Sci., 1*(2), 89-97.

Conant, R. C., & Ross Ashby, W. (1970). Every good regulator of a system must be a model of that system. *International journal of systems science, 1*(2), 89-97.

Constant, A., Bervoets, J., Hens, K., & Van de Cruys, S. (2018). Precise worlds for certain minds: An ecological perspective on the relational self in autism. *Topoi*.

Constant, A., Ramstead, M., Veissière, S., Campbell, J., & Friston, K. (2018). A variational approach to niche construction. *Journal of the Royal Society Interface*.

Constant, A., Ramstead, M. J., Veissière, S. P., & Friston, K. (2019). Regimes of Expectations: An Active Inference Model of Social Conformity and Decision Making. *Frontiers in Psychology, 10*, 679.

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences, 5*(12), 539-546.

Corlett, P. R., & Fletcher, P. C. (2014). Computational psychiatry: a Rosetta Stone linking the brain to mental illness. *The Lancet Psychiatry, 1*(5), 399-402.

Corlett, P. R., Frith, C. D., & Fletcher, P. C. (2009). From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology-Berlin, 206*(4), 515-530.

Costall, A. (1997). The meaning of things. *Social Analysis: The International Journal of Social and Cultural Practice, 41*(1), 76-85.

Crauel, H., & Flandoli, F. (1994). Attractors for random dynamical systems. *Probab Theory Relat Fields, 100*, 365-393.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*: Oxford University Press.

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology & Philosophy, 22*(4), 547-563.

Cummins, R. (1983). The nature of psychological explanation. In. Cambridge, MA: MIT Press.

Cziko, G. (1995). *Without miracles: universal selection theory and the second Darwinian revolution*. Cambridge, MA: MIT press.

Daunizeau, J. (2018). A plea for "variational neuroethology": Comment on "Answering Schrödinger's question: A free-energy formulation" by MJ Desormeau Ramstead et al. *Physics of life Reviews*.

Dauwels, J. (2007, 24-29 June 2007). *On Variational Message Passing on Factor Graphs.* Paper presented at the 2007 IEEE International Symposium on Information Theory.

Davidson, D. (1980/1970). Mental events. In *Actions and events* (pp. 207-225). Oxford: Clarendon Press.

Davis, M. J. (2006). Low-dimensional manifolds in reaction- diffusion equations. 1. Fundamental aspects. *The Journal of Physical Chemistry A, 110*(16), 5235-5256.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience, 8*(12), 1704.

Dawkins, R. (1982). *The Extended Phenotype: The Long Reach of the Gene*. Oxford: Oxford University Press.

Dawson, M. R. (2013). *Mind, body, world: Foundations of cognitive science*. Edmonton: Athabasca University Press.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The helmholtz machine. *Neural computation, 7*(5), 889-904.

De Haan, S., Rietveld, E., Stokhof, M., & Denys, D. (2013). The phenomenology of deep brain stimulation-induced changes in OCD: An enactive affordance-based model. *Frontiers in human neuroscience, 7*(653), 1-14.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences, 6*(4), 485-507.

den Ouden, H. E., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *Journal of Neuroscience, 30*(9), 3210-3219.

Dennett, D. C. (1991). Real patterns. *The Journal of philosophy, 88*(1), 27-51.

Depew, D. J., & Weber, B. H. (1995). *Darwinism evolving: Systems dynamics and the genealogy of natural selection*. Cambridge, MA: MIT Press.

Di Paolo, E. (2009). Extended life. *Topoi, 28*(1), 9.

Di Paolo, E., Buhrmann, T., & Barandiaran, X. (2017). *Sensorimotor life: An enactive proposal*: Oxford University Press.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences, 4*(4), 429-452.

Di Paolo, E. A., & Thompson, E. (2014). The enactive approach. In L. E. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 68-78). London: Routledge.

Dissanayake, E. (2009). The artification hypothesis and its relevance to cognitive science, evolutionary aesthetics, and neuroaesthetics. *Cognitive Semiotics, 5*(fall2009), 136-191.

Dorogovtsev, S. N., & Mendes, J. F. (2002). Evolution of networks. *Advances in physics, 51*(4), 1079-1187.

Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology, 53*, 109-132.

Eigen, M., & Schuster, P. (1979). *The hypercycle: a principle of natural self-organisation*. Berlin: Springer-Verlag.

Elfwing, S., Uchibe, E., & Doya, K. (2016). From free energy to expected energy: Improving energy-based value function approximation in reinforcement learning. *Neural Networks: The Official Journal of the International Neural Network Society, 84*(17-27).

Eliasmith, C. (2005). A new perspective on representational problems. *Journal of Cognitive Science, 6*(97), 123.

Engel, A. K., Friston, K. J., & Kragic, D. (2016). The pragmatic turn: Toward action-oriented views in cognitive science.

England, J. L. (2013). Statistical physics of self-replication. *J Chem Phys, 139*(12), 121923. doi:10.1063/1.4818538

Evans, D. J., & Searles, D. J. (2002). The Fluctuation Theorem. *Advances in physics, 51*(7), 1529-1585.

Fabry, R. E. (2017). Betwixt and between: the enculturated predictive processing approach to cognition. *Synthese*, 1-36.

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in human neuroscience, 4*, 215.

Feynman, R. (1972). *Statistical mechanics: a set of lectures*. Reading, MA: Benjamin/Cummings Publishing.

Finlay, B. L., & Uchiyama, R. (2015). Developmental mechanisms channeling cortical evolution. *Trends in neurosciences, 38*(2), 69-76.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection* Oxford Clarendon Press

Fodor, J. A. (1975). *The language of thought* (Vol. 5): Harvard University Press.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT press.

Frank, S. A. (2012). Natural selection. V. How to read the fundamental equations of evolutionary change in terms of information theory. *Journal of evolutionary biology, 25*(12), 2377-2396.

Frank, T. D. (2004). *Nonlinear Fokker-Planck Equations: Fundamentals and Applications*. Berlin: Springer.

Frankenhuis, W. E., Panchanathan, K., & Clark Barrett, H. (2013). Bridging developmental systems theory and evolutionary psychology using dynamic optimization. *Developmental Science, 16*(4), 584-598.

Freeman, W. J. (1994). Characterization of state transitions in spatially distributed, chaotic, nonlinear, dynamical systems in cerebral cortex. *Integr Physiol Behav Sci., 29*(3), 294-306.

Frijda, N. H. (1986). *The emotions*. Cambridge: Cambridge University Press.

Frijda, N. H. (2007). *The laws of emotions*. Mahwah, NJ: Erlbaum.

Friston, K. (2010a). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127-138.

Friston, K. (2011a). Embodied inference: or ‘‘I think therefore I am, if I am what I think’’. In W. Tschacher & C. Bergomi (Eds.), *The implications of embodiment: Cognition and communication* (pp. 89-125). Exeter, UK: Imprint Academic.

Friston, K. (2012a). A free energy principle for biological systems. *Entropy, 14*(11), 2100-2121.

Friston, K. (2012b). Predictive coding, precision and synchrony. *Cognitive neuroscience, 3*(3-4), 238-239.

Friston, K. (2013a). Active inference and free energy. *Behavioral and Brain Sciences, 36*(3), 212-213.

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

Friston, K. (in preparation). *A free energy principle for a particular physics*.

Friston, K., & Ao, P. (2011). Free energy, value, and attractors. *Computational and mathematical methods in medicine, 2012*.

Friston, K., & Buzsáki, G. (2016). The functional anatomy of time: what and when in the brain. *Trends in Cognitive Sciences, 20*(7), 500-511.

Friston, K., Daunizeau, J., & Kiebel, S. (2009). Reinforcement learning or active inference? *PloS one, 4*(7), e6421.

Friston, K., Daunizeau, J., Kilner, J., & Kiebel, S. (2010a). Action and behavior: a free-energy formulation. *Biological cybernetics, 102*(3), 227-260.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2016a). Active inference and learning. *Neuroscience & Biobehavioral Reviews, 68*, 862-879.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active Inference: A Process Theory. *Neural Comput, 29*(1), 1-49. doi:10.1162/NECO_a_00912

Friston, K., & Frith, C. (2015a). A duet for one. *Consciousness and cognition, 36*, 390-405.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical transactions of the Royal Society B: Biological sciences, 364*(1521), 1211-1221.

Friston, K., Kilner, J., & Harrison, L. (2006a). A free energy principle for the brain. *J Physiol Paris, 100*(1-3), 70-87. doi:10.1016/j.jphysparis.2006.10.001

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological cybernetics, 104*(1-2), 137-160.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive neuroscience, 6*(4), 187-214.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Human brain mapping, 2*(1-2), 56-78.

Friston, K. J. (2005a). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences, 360*(1456), 815-836.

Friston, K. J. (2005b). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360*(1456), 815-836.

Friston, K. J. (2010b). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127-138.

Friston, K. J. (2011b). Functional and effective connectivity: a review. *Brain connectivity, 1*(1), 13-36.

Friston, K. J., Bastos, A. M., Pinotsis, D., & Litvak, V. (2015). LFP and oscillations—what do they tell us? *Current opinion in neurobiology, 31*, 1-6.

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010b). Action and behavior: a free-energy formulation. *Biol Cybern., 102*(3), 227-260.

Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2016b). Active inference: a process theory. *Neural computation, 29*, 1-49.

Friston, K. J., & Frith, C. D. (2015b). Active inference, communication and hermeneutics. *cortex, 68*, 129-143.

Friston, K. J., Kilner, J., & Harrison, L. (2006b). A free energy principle for the brain. *Journal of Physiology-Paris, 100*(1), 70-87.

Friston, K. J., Levin, M., Sengupta, B., & Pezzulo, G. (2015). Knowing one's place: a free-energy approach to pattern regulation. *Journal of the Royal Society Interface, 12*(105).

Friston, K. J., Parr, T., & de Vries, B. (2017). The graphical brain: belief propagation and active inference. *Network Neuroscience, 1*(4), 381-414.

Friston, K. J., Redish, A. D., & Gordon, J. A. (2017). Computational nosology and precision psychiatry. *Computational Psychiatry, 1*, 2-23.

Friston, K. J., Rosch, R., Parr, T., Price, C., & Bowman, H. (2018). Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews, 90*, 486-501.

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., . . . Bestmann, S. (2012). Dopamine, affordance and active inference. *PLOS Computational Biology, 8*(1), e1002327.

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese, 159*(3), 417-458.

Friston, K. J., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. *The Lancet Psychiatry, 1*(2), 148-158.

Frith, C. (2007). *Making up the mind: How the brain creates our mental world*. Malden, MA: Blackwell John Wiley & Sons.

Frith, C. D., & Friston, K. J. (2013). False perceptions and false beliefs: understanding schizophrenia. *Neurosciences and the human person: New perspectives on human activities. Pontifical Academy of Sciences, Scripta Varia, 121*, 1-15.

Froese, T., & Di Paolo, E. A. (2011). The enactive approach: Theoretical sketches from cell to society. *Pragmatics & Cognition, 19*(1), 1-36.

Fuchs, T., & De Jaegher, H. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. . *Phenomenology and the Cognitive Sciences, 8*(4), 465-486.

Fuentes, A. (2014). Human evolution, niche complexity, and the emergence of a distinctively human imagination. *Time and mind, 7*(3), 241-257.

Gadamer, H.-G. (2010/1927). *Truth and Method* (J. Weinsheimer & D. G. Marshall, Trans.). London and New York: Continuum.

Gallagher, S. (2001). The practice of mind. Theory, simulation or primary interaction? *Journal of Consciousness Studies, 8*(5-6), 83-108.

Gallagher, S. (2006). *How the body shapes the mind*: Clarendon Press.

Gallagher, S. (2008). Inference or interaction: social cognition without precursors. *Philosophical Explorations, 11*(3), 163-174.

Gallagher, S. (2012). On the possibility of naturalizing phenomenology. *Oxford Handbook of Contemporary Phenomenology, 4*, 70-93.

Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*: Oxford University Press.

Gallagher, S., & Allen, M. (2016). Active inference, enactivism and the hermeneutics of social cognition. *Synthese*, 1-22.

Gallagher, S., & Ransom, T. G. (2016). Artifacting minds: Material engagement theory and joint action. *Embodiment in evolution and culture*, 337-351.

Gibson, J. J. (1979). *The ecological approach to visual perception*: Psychology Press.

Gładziejewski, P. (2016). Predictive coding and representationalism. *Synthese, 193*(2), 559-582.

Gładziejewski, P., & Miłkowski, M. (2017). Structural representations: causally relevant and different from detectors. *Biology & Philosophy*, 1-19.

Goffman, E. (1971). *Relations in Public: Microstudies of the Public Order*. New York, NY: Basic Books.

Gold, I. (2009). Reduction in psychiatry. *The Canadian Journal of Psychiatry, 54*(8), 506-512.

Gold, I., & Kirmayer, L. J. (2007). Cultural psychiatry on Wakefield's procrustean bed. *World Psychiatry, 6*(3), 165.

Goldstone, R. L., Landy, D., & Brunel, L. C. (2011). Improving perception to make distant connections closer. *Frontiers in Psychology, 2*(385). doi:doi:10.3389/fpsyg.2011.00385

Grice, H. (1989a). *Study in the way of words*. Cambridge, MA: Harvard University Press.

Grice, H. P. (1957). Meaning. *The philosophical review, 66*(3), 377-388.

Grice, H. P. (1969). Utterer's meaning and intention. *The philosophical review, 78(2)*(2), 147-177.

Grice, H. P. (1971). Intention and uncertainty. *Oxford, Oxford University Press*.

Grice, H. P. (1989b). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Grush, R. (2001). The semantic challenge to computational neuroscience. In P. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and method in the neurosciences* (pp. 155-172). Pittsburgh: University of Pittsburgh Press.

Gu, S., Satterthwaite, T. D., Medaglia, J. D., Yang, M., Gur, R. E., Gur, R. C., & Bassett, D. S. (2015). Emergence of system roles in normative neurodevelopment. *Proceedings of the National Academy of Sciences, 112*(44), 13681-13686.

Guze, S. B. (1989). Biological psychiatry: is there any other kind? *Psychological Medicine, 19*(2), 315-323.

Hacking, I. (1995). The looping effect of human kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal Cognition: A Multidisciplinary Debate* (pp. 351-383). Oxford: Oxford University Press.

Hacking, I. (1999). *The social construction of what?* : Harvard university press.

Hacking, I. (2002). *Historical Ontology*. Cambridge, MA: Harvard University Press.

Hacking, I. (2004). Between Michel Foucault and Erving Goffman: between discourse in the abstract and face-to-face interaction. *Economy and society, 33*(3), 277-302.

Haken, H. (1977). Synergetics. *Physics Bulletin, 28*(9), 412.

Haken, H. (1983). *Synergetics: An introduction. Non-equilibrium phase transition and self-selforganisation in physics, chemistry and biology*. Berlin: Springer-Verlag.

Haken, H. (1996). *Principles of brain functioning: a synergetic approach to brain activity, behaviour and cognition*. Berlin: Springer-Verlag.

Haken, H. (2013). *Synergetics: introduction and advanced topics*: Springer Science & Business Media.

Harper, M. (2011). Escort evolutionary game theory. *Physica D: Nonlinear Phenomena, 240*(18), 1411-1415.

Haugeland, J. (1990). The intentionality all-stars. *Philosophical perspectives, 4*, 383-427.

Heft, H. (2001). *Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James's radical empiricism*: Psychology Press.

Heidegger, M. (2010/1927). *Being and time*: SUNY Press.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science, 15*(2), 135-175.

Henrich, J. (2015). *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton, NJ: Princeton University Press.

Henriques, G. (2011). *A new unified theory of psychology*. New York, NY: Springer

Heras-Escribano, M. (2019a). *The Philosophy of Affordances* Palgrave Macmillan.

Heras-Escribano, M. (2019b). Pragmatism, enactivism, and ecological psychology: towards a unified approach to post-cognitivism. *Synthese*, 1-27.

Hesp, C., Ramstead, M., Constant, A., Badcock, P. B., Kirchhoff, M., & Friston, K. (2019). A multi-scale view of the emergent complexity of life: A free-energy proposal. In M.

P. e. al. (Ed.), *Evolution, Development, and Complexity: Multiscale Models in Complex Adaptive Systems*: Springer.

Hesse, J., & Gross, T. (2014). Self-organized criticality as a fundamental property of neural systems. *Frontiers in Systems Neuroscience, 8*(166).

Heyes, C. (2012). New thinking: the evolution of human cognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 367*(1599), 2091-2096.

Heyes, C. (2018). *Cognitive gadgets: the cultural evolution of thinking*: Harvard University Press.

Hilgetag, C. C., & Hütt, M.-T. (2014). Hierarchical modular brain connectivity is a stretch for criticality. *Trends in Cognitive Sciences, 18*(3), 114-115.

Hinton, G. E., & Zemel, R. S. (1993). *Autoencoders, minimum description length and Helmholtz free energy*. Paper presented at the Proceedings of the 6th International Conference on Neural Information Processing Systems, Denver, Colorado.

Hirschfeld, L. A. (1998). *Race in the making: Cognition, culture, and the child's construction of human kinds*: MIT Press.

Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: a framework for understanding uncertainty-related anxiety. *Psychological review, 119*(2), 304-320.

Hohwy, J. (2014). *The predictive mind*. Oxford: Oxford University Press.

Hohwy, J. (2016). The self-evidencing brain. *Noûs, 50*(2), 259-285.

Hohwy, J. (2017). How to Entrain Your Evil Demon. *Philosophy and predictive processing. Frankfurt am Main: MIND Group*.

Hohwy, J., Roepstorff, A., & Friston, K. J. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition, 108*(3), 687-701.

Howes, D. (2011). Reply to Tim Ingold. *Social Anthropology, 19*(3), 318-322.

Hrdy, S. B. (2011). *Mothers and others*: Harvard University Press.

Hristovski, R., Davids, K., Araújo, D., & Button, C. (2006). How boxers decide to punch a target: emergent behaviour in nonlinear dynamical movement systems. *Journal of sports science & medicine, 5*(CSSI), 60.

Hristovski, R., Davids, K. W., & Araujo, D. (2009). Information for regulating action in sport: metastability and emergence of tactical solutions under ecological constraints. In *Perspectives on cognition and action in sport* (pp. 43-57): Nova Science Publishers, Inc.

Huang, G. T. (2008). Essence of thought. *New Scientist, 198*(2658), 30-33.

Huneman, P., & Machery, E. (2015). Evolutionary psychology: issues, results, debates. In *Handbook of Evolutionary Thinking in the Sciences* (pp. 647-657): Springer.

Husserl, E. (1990). *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy: Second book studies in the phenomenology of constitution* (Vol. 3): Springer Science & Business Media.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT press.

Hutchins, E. (2010). Cognitive ecology. *Topics in cognitive science, 2*(4), 705-715.

Hutchins, E. (2014). The cultural ecosystem of human cognition. *Philosophical Psychology, 27*(1), 34-49.

Hutto, D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content*. Cambridge, MA: MIT Press.

Hutto, D., & Satne, G. (2015). The natural origins of content. *Philosophia, 43*(3), 521-536.

Hutto, D. D., Kirchhoff, M. D., & Myin, E. (2014). Extensive enactivism: why keep it all in? *Frontiers in human neuroscience, 8*, 706.

Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. Cambridge, MA: MIT Press.

Ingold, T. (2000). *The perception of the environment essays on livelihood, dwelling and skill*. New York: Routledge.

Ingold, T. (2001a). From the transmission of representations to the education of attention. *The debated mind: Evolutionary psychology versus ethnography*, 113-153.

Ingold, T. (2001b). From the transmission of representations to the education of attention. In H. Whitehouse (Ed.), *The debated mind: Evolutionary psychology versus ethnography* (pp. 113-153). Oxford: Berg Publishers.

Insel, T. R., & Quirion, R. (2005). Psychiatry as a clinical neuroscience discipline. *Jama, 294*(17), 2221-2224.

James, W. (1975). *Pragmatism* (H. Thayer Ed.): Harvard University Press.

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters, 78*(14), 2690-2693.

Joffily, M., & Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLOS Computational Biology, 9*(6).

Juarrero, A. (1999). *Dynamics in action*. Cambridge, MA: MIT Press.

Jung, M., Hwang, J., & Tani, J. (2015). Self-organization of spatio-temporal hierarchy via learning of dynamic visual image patterns on action sequences. *PloS one, 10*(7).

Kaiser, M., Hilgetag, C. C., & Kötter, R. (2010). Hierarchy and dynamics of neural networks. *Frontiers in Neuroinformatics, 4*(112), 4-6.

Kaplan, R., & Friston, K. J. (2018). Planning and navigation as active inference. *Biological cybernetics*, 1-21.

Kauffman, S. A. (1993). *The origins of order: self-organization and selection in evolution*. Oxford: Oxford University Press.

Keane, W. (2014). Affordances and reflexivity in ethical life: An ethnographic stance. *Anthropological Theory, 14*(1), 3-26.

Kelly, D., Faucher, L., & Machery, E. (2010). Getting rid of racism: Assessing three proposals in light of psychological evidence. *Journal of Social Philosophy, 41*(3), 293-322.

Kelso, J. S. (1995). *Dynamic patterns: the self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Kendler, K., & Parnas, J. (2008). *Philosophical issues in psychiatry: Explanation, nosology and phenomenology*. Baltimore, MD: Johns Hopkins University Press.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008a). A hierarchy of time-scales and the brain. *PLOS Computational Biology, 4*(11).

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008b). A hierarchy of time-scales and the brain. *PLOS Computational Biology, 4*(11), e1000209.

Kiebel, S. J., & Friston, K. J. (2011). Free energy and dendritic self-organization. *Frontiers in Systems Neuroscience, 5*.

Kiefer, A., & Hohwy, J. (2017). Content and misrepresentation in hierarchical generative models. *Synthese*, 1-29.

Kim, J. (1989). Mechanism, purpose, and explanatory exclusion. *Philosophical perspectives, 3*, 77-108.

Kim, J. (1993). *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.

Kim, J. (1999). Making sense of emergence. *Philosophical Studies, 95*(1), 3-36.

Kim, J. (2006). Emergence: Core ideas and issues. *Synthese, 151*(3), 547-559.

Kinzler, K. D., & Spelke, E. S. (2011). Do infants show social preferences for people differing in race? *Cognition, 119*(1), 1-9.

Kirchhoff, M. (2015a). Experiential fantasies, prediction, and enactive minds. *Journal of Consciousness Studies, 22*(3-4), 68-92.

Kirchhoff, M. (2015b). Species of realization and the free energy principle. *Australasian Journal of Philosophy, 93*(4), 706-723.

Kirchhoff, M. (2016). Autopoiesis, free energy, and the life–mind continuity thesis. *Synthese*, 1-22. doi:doi:10.1007/s11229-016-1100-6

Kirchhoff, M. (2018a). Autopoiesis, free energy, and the life–mind continuity thesis. *Synthese, 195*(6), 2519-2540.

Kirchhoff, M. (2018b). Hierarchical Markov Blankets and Adaptive Active Inference : Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Kirchhoff, M. (2018c). Predictive processing, perceiving and imagining: Is to perceive to imagine, or something close to it? *Philosophical Studies, 175*(3), 751-767.

Kirchhoff, M., & 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., & Kiverstein, J. (2019). *Extended Consciousness and Predictive Processing: A Third-Wave View*. New York Routledge.

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface, 15*(138), 20170792.

Kirchhoff, M., & Robertson, I. (2018). Enactivism and predictive processing: a non-representational view. *Philosophical Explorations, 21*(2), 264-281.

Kirchhoff, M. D. (2012). Extended cognition and fixed properties: steps to a third-wave version of extended cognition. *Phenomenology and the Cognitive Sciences, 11*(2), 287-308.

Kirmayer, L., & Ramstead, M. (2017). Embodiment and Enactment in Cultural Psychiatry. In *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*: MIT Press

Kirmayer, L. J. (2008). Culture and the metaphoric mediation of pain. *Transcultural Psychiatry, 45*(2), 318-338.

Kirmayer, L. J. (2015). Re-visioning psychiatry: Toward an ecology of mind in health and illness. In L. J. Kirmayer, R. Lemelson, & C. Cummings (Eds.), *Re-visioning psychiatry: cultural phenomenology, critical neuroscience and global mental health* (pp. 622-660). New York: Cambridge University Press.

Kirmayer, L. J. (2018). Ontologies of life: From thermodynamics to teleonomics. Comment on “Answering Schrödinger’s question: A free-energy formulation” by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Kirmayer, L. J., & Bhugra, D. (2009). Culture and mental illness: social context and explanatory models. *Psychiatric diagnosis: patterns and prospects. New York: John Wiley & Sons*, 29-37.

Kirmayer, L. J., & Gold, I. (2012). Re-socializing psychiatry. *Critical neuroscience: A handbook of the social and cultural contexts of neuroscience*.

Kirmayer, L. J., Lemelson, R., & Cummings, C. A. (2015). *Re-visioning psychiatry: Cultural phenomenology, critical neuroscience, and global mental health*: Cambridge University Press.

Kistler, M. (2009). Mechanisms and downward causation. *Philosophical Psychology, 22*(5), 595-609.

Kitano, H. (2002). Computational systems biology. *Nature, 420*(6912), 206.

Kiverstein, J., & Clark, A. (2009). Introduction: Mind embodied, embedded, enacted: One church or many? *Topoi, 28*(1), 1-7.

Kiverstein, J., & Rietveld, E. (2013). Dealing with context through action-oriented predictive processing. *Frontiers in Psychology, 3*, 421.

Kiverstein, J., & Rietveld, E. (2015). The primacy of skilled intentionality: on Hutto & Satne’s the natural origins of content. *Philosophia, 43*(3), 701-721.

Kiverstein, J., & Wheeler, M. (2012). *Heidegger and cognitive science*: Palgrave Macmillan.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in neurosciences, 27*(12), 712-719.

Kok, P., Brouwer, G. J., van Gerven, M. A., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience, 33*(41), 16275-16284.

Kok, P., Jehee, J. F., & De Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron, 75*(2), 265-270.

Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*: University of Chicago press.

Laland, K., Matthews, B., & Feldman, M. W. (2016). An introduction to niche construction theory. *Evolutionary Ecology, 30*, 191-202.

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in human neuroscience, 302*.

Levin, S. (1998). Ecosystems and the biosphere as complex adaptive systems. *Ecosystems, 1*(5), 431-436.

Levin, S. (2003). Complex adaptive systems: exploring the known, the unknown and the unknowable. *Bulletin of the American Mathematical Society, 40*(1), 3-19.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT press.

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Blackwell.

Lewontin, R. C. (1982). Organism and environment. In H. C. Plotkin (Ed.), *Learning, development, and culture: Essays in evolutionary epistemology*. New York: Wiley.

Leydesdorff, L. (1995). *The challenge of scientometrics: The development, measurement, and self-organization of scientific communications*. Leiden: DSWO Press.

Leydesdorff, L. (2001). *A sociological theory of communication: The self-organization of the knowledge-based society*. Parkland, FL: Universal Publishers.

Leydesdorff, L. (2018). Lifting the Markov blankets of socio-cultural evolution: A comment on “Answering Schrödinger's question: A free-energy formulation” by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Linson, A., Clark, A., Ramamoorthy, S., & Friston, K. (2018). The active inference approach to ecological perception: general information dynamics for natural and artificial embodied cognition. *Frontiers in Robotics and AI, 5*, 21.

Machery, E., & Faucher, L. (2005). Why do we think racially? In *Handbook of categorization in cognitive science* (pp. 1009-1033): Elsevier.

MacKay, D. J. (1995). Free-energy minimisation algorithm for decoding and cryptoanalysis. *Electronics Letters, 31*, 445-447.

Mantegna, R. N., & Stanley, H. E. (1995). Scaling behaviour in the dynamics of an economic index. *Nature, 376*(6535), 46-49.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information, henry holt and co*. San Francisco: W. H. Freeman.

Martyushev, L., & Seleznev, V. (2006). Maximum entropy production principle in physics, chemistry and biology. *Physics reports, 426*(1), 1-45.

Martyushev, L. M. (2018). Living systems do not minimize free energy: Comment on “Answering Schrödinger's question: A free-energy formulation” by Maxwell James Dèsormeau Ramstead et al. *Physics of life Reviews*.

Maturana, H. R., & Varela, F. J. (1980). Autopoiesis: The organization of the living. *Autopoiesis and cognition: The realization of the living, 42*, 59-138.

McLaughlin, B. P. (1992). The rise and fall of British emergentism. *Emergence or reduction*, 49-93.

Mead, M. (1975). *Growing up in New Guinea: A comparative study of primitive education*. New York: Morrow.

Meltzoff, A. N., & Prinz, W. (Eds.). (2002). *The imitative mind: Development, evolution and brain bases*. Cambridge: Cambridge University Press.

Menary, R. (2010). The Extended Mind and Cognitive Integration. In R. Menary (Ed.), *The extended mind*. Cambridge, MA: MIT Press

Mengistu, H., Huizinga, J., Mouret, J.-B., & Clune, J. (2016). The evolutionary origins of hierarchy. *PLOS Computational Biology, 12*(6).

Merleau-Ponty, M. (1968/1964). *The visible and the invisible* (A. Lingus, Trans.). Evanston, IL: Northwestern University Press.

Merleau-Ponty, M. (2013/1945). *Phenomenology of perception*: Routledge.

Mesoudi, A., Whiten, A., & Laland, K. N. (2006). Towards a unified science of cultural evolution. *Behavioral and Brain Sciences, 29*(04), 329-347.

Metzinger, T., & Wiese, W. (Eds.). (2017). *Philosophy and predictive processing*. Frankfurt am Main: MIND Group.

Michael, J., Christensen, W., & Overgaard, S. (2014). Mindreading as social expertise. *Synthese, 191*(5), 817-840.

Miłkowski, M. (2013). *Explaining the computational mind*. Cambridge, MA: MIT Press.

Mill, J. S. (1843). *A system of logic*: Longmans, Green, Reader, and Dryer.

Miller, J. H., & Page, S. E. (2009). *Complex adaptive systems: an introduction to computational models of social life*. Princeton, NJ: Princeton University Press.

Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*: MIT press.

Millikan, R. G. (2004). *Varieties of meaning*. Cambridge, MA: MIT press.

Millikan, R. G. (2005). *Language: A biological model*: Oxford University Press on Demand.

Moore, R. (2013). Social learning and teaching in chimpanzees. *Biology & Philosophy, 28*(6), 879-901.

Morgan, C. L. (2013/1923). *Emergent evolution*: Read Books Ltd.

Moya, C., & Henrich, J. (2016). Culture–gene coevolutionary psychology: cultural learning, language, and ethnic psychology. *Current Opinion in Psychology, 8*, 112-118.

Murphy, D. (2010). Explanation in psychiatry. *Philosophy Compass, 5*(7), 602-610.

Newen, A., De Bruin, L., & Gallagher, S. (2018). *The Oxford handbook of 4E cognition*: Oxford University Press.

Nicolis, G., & Prigogine, I. (1977). *Self-organization in nonequilibrium systems*. New York, NY: John Wiley.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological bulletin, 131*(4), 510-532.

Noë, A. (2004). *Action in perception*: MIT press.

O'Brien, G., & Opie, J. (2004). Notes toward a structuralist theory of mental representation. In *Representation in mind* (pp. 1-20): Elsevier.

O'Brien, G., & Opie, J. (2004). Notes toward a structuralist theory of mental representation. In H. Clapin, P. Staines, & P. Slezak (Eds.), *Representation in mind: New approaches to mental representation* (pp. 1-20).

O'Brien, G., & Opie, J. (2009). The role of representation in computation. *Cognitive processing, 10*(1), 53-62.

O'Brien, G., & Opie, J. (2015). Intentionality lite or analog content? *Philosophia, 43*(3), 723-729.

Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). *Niche construction: The neglected process in evolution*. Princeton: Princeton University Press.

Odling-Smee, J. (1988). Niche constructing phenotypes. In H. Plotkin (Ed.), *The role of behavior in evolution*. Cambridge, MA: MIT Press.

Odling-Smee, J., & Laland, K. N. (2011). Ecological inheritance and cultural inheritance: what are they and how do they differ? *Biological Theory, 6*(3), 220-230.

Odum, H. T. (1994). *Ecological and general systems: an introduction to systems ecology*: Univ. Press of Colorado.

Oudeyer, P.-Y., & Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in neurorobotics, 1*(6).

Palacios, E., Razi, A., Parr, T., Kirchhoff, M., & Friston, K. (2017). Biological self-organisation and Markov blankets. *bioRxiv*, 227181.

Park, H.-J., & Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science, 342*(6158).

Parr, T., & Friston, K. J. (2017a). Uncertainty, epistemics and active inference. *Journal of the Royal Society Interface, 14*(136), 20170376.

Parr, T., & Friston, K. J. (2017b). Working memory, attention, and salience in active inference. *Scientific reports, 7*(1), 14678.

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

Pauker, K., Williams, A., & Steele, J. (2016). Children's racial categorization in context. *Child development perspectives, 10*(1), 33-38.

Paul, C. (2006). Morphological computation: A basis for the analysis of morphology and control requirements. *Robotics and Autonomous Systems, 54*(8), 619-630.

Pearl, J. (1988). Probabilistic reasoning in intelligent systems: Networks of plausible inference. In. San Mateo, CA: Morgan Kaufmann.

Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*: Basic Books.

Perdikis, D., Huys, R., & Jirsa, V. K. (2011). Time scale hierarchies in the functional organization of complex behaviors. *PLOS Computational Biology, 7*(9), e1002198.

Petitot, J., Varela, F. J., Pachoud, B., & Roy, J.-M. (Eds.). (1999). *Naturalizing phenomenology: Issues in contemporary phenomenology and cognitive science*: Stanford University Press.

Pezzulo, G., & 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., & Levin, M. (2018). Embodying Markov blankets. Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Pfeifer, R., & Gómez, G. (2009). Morphological computation–connecting brain, body, and environment. In *Creating brain-like intelligence* (pp. 66-83): Springer.

Pfeifer, R., Iida, F., & Gómez, G. (2006). *Morphological computation for adaptive behavior and cognition.* Paper presented at the International Congress Series.

Piccinini, G. (2015). *Physical computation: A mechanistic account*: OUP Oxford.

Piccinini, G., & Scarantino, A. (2011). Information processing, computation, and cognition. *Journal of biological physics, 37*(1), 1-38.

Poldrack, R. A. (2010). Mapping mental function to brain structure: how can cognitive neuroimaging succeed? *Perspectives on Psychological Science, 5*(6), 753-761.

Price, C. J., & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology, 22*(3-4), 262-275.

Prigogine, I., & Stengers, I. (1984). *Order out of chaos: man's new dialogue with nature*. New York, NY: Bantam Books

Putnam, H. (1975). *Mind, language, and reality*. Cambridge: Cambridge University Press.

Queller, D. C. (2014). Joint phenotypes, evolutionary conflict and the fundamental theorem of natural selection. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 369*(1642), 20130423.

Quine, W. (1961). *From a Logical Point of View*. Cambridge, MA Harvard University Press

Ramstead, M., Badcock, P. B., & Friston, K. (2018a). Answering Schrödinger's question: A free-energy formulation. *Physics of life Reviews, 24*, 1-16.

Ramstead, M., Badcock, P. B., & Friston, K. (2018b). Variational neuroethology: Answering further questions: Reply to comments on "Answering Schrödinger's question: A free-energy formulation". *Physics of life Reviews, 24*, 59-66.

Ramstead, M., Constant, A., Badcock, P. B., & Friston, K. (2019). Variational ecology and the physics of sentient systems. *Physics of life Reviews*.

Ramstead, M., Veissière, S., & Kirmayer, L. (2016). Cultural affordances: scaffolding local worlds through shared intentionality and regimes of attention. *Frontiers in Psychology, 7*.

Ramstead, M. J. (2015). Naturalizing what? Varieties of naturalism and transcendental phenomenology. *Phenomenology and the Cognitive Sciences, 14*(4), 929-971.

Ramstead, M. J., Kirchhoff, M. D., Constant, A., & Friston, K. J. (2019). Multiscale integration: Beyond internalism and externalism. *Synthese*.

Rao, R. P., & 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.

Redish, A. D., & Gordon, J. A. (Eds.). (2016). *Computational psychiatry: New perspectives on mental illness*. Cambridge, MA: MIT Press

Richeson, J. A., & Sommers, S. R. (2016). Toward a social psychology of race and race relations for the twenty-first century. *Annual review of psychology, 67*, 439-463.

Rietveld, E. (2008a). Situated normativity: The normative aspect of embodied cognition in unreflective action. *Mind, 117*, 973-1001.

Rietveld, E. (2008b). Special section: The skillful body as a concernful system of possible actions phenomena and neurodynamics. *Theory & Psychology, 18*, 341-363.

Rietveld, E. (2008c). *Unreflective action. A philosophical contribution to integrative neuroscience*. Amsterdam: Institute for Logic, Language and Computation.

Rietveld, E. (2012). Bodily intentionality and social affordances in context. In F. Paglieri (Ed.), *Consciousness in interaction. The role of the natural and social context in shaping consciousness* (pp. 207-226). Amsterdam: J. Benjamins.

Rietveld, E., De Haan, S., & Denys, D. (2013). Social affordances in context: What is it that we are bodily responsive to? *Behavioral and Brain Sciences, 36*(4), 436.

Rietveld, E., & Kiverstein, J. (2014). A rich landscape of affordances. *Ecological Psychology, 26*(4), 325-352.

Robbins, J., & Rumsey, A. (2008). Introduction: Cultural and linguistic anthropology and the opacity of other minds. *Anthropological Quarterly, 81*(2), 407-420.

Roepstorff, A. (2013). Interactively human: Sharing time, constructing materiality. *Behavioral and Brain Sciences, 36*(3), 224-225.

Roepstorff, A., & Frith, C. (2004). What's at the top in the top-down control of action? Script-sharing and 'top-top'control of action in cognitive experiments. *Psychological Research, 68*(2-3), 189-198.

Roepstorff, A., Niewöhner, J., & Beck, S. (2010). Enculturing brains through patterned practices. *Neural Networks, 23*(8), 1051-1059.

Rogoff, B. (2003). *The cultural nature of human development*: Oxford university press.

Rumsey, A. (2013). Intersubjectivity, deception and the 'opacity of other minds': Perspectives from Highland New Guinea and beyond. *Language & Communication, 33*(3), 326-343.

Rupert, R. D. (2004). Challenges to the hypothesis of extended cognition. *The Journal of philosophy, 101*(8), 389-428.

Rupert, R. D. (2009). *Cognitive systems and the extended mind*: Oxford University Press.

Sacheli, L. M., Christensen, A., Giese, M. A., Taubert, N., Pavone, E. F., Aglioti, S. M., & Candidi, M. (2015). Prejudiced interactions: implicit racial bias reduces predictive simulation during joint action with an out-group avatar. *Scientific reports, 5*, 8507.

Salge, C., Glackin, C., & Polani, D. (2014). Changing the environment based on empowerment as intrinsic motivation. *Entropy, 16*(5), 2789-2819.

Sarkar, S. (Ed.) (1992). *The Founders of Evolutionary Genetics: A Centenary Reappraisal* (Vol. 142). Dordrecht / Boston / London: Kluwer Academic Publishers

Satne, G. (2015). The social roots of normativity. *Phenomenology and the Cognitive Sciences, 14*(4), 673-682.

Scarantino, A. (2015). Information as a probabilistic difference maker. *Australasian Journal of Philosophy, 93*(3), 419-443.

Scarantino, A., & Piccinini, G. (2010). Information without truth. *Metaphilosophy, 41*(3), 313-330.

Schaffner, K. F. (1993). *Discovery and explanation in biology and medicine*: University of Chicago press.

Schieffelin, B. B., & Ochs, E. (1986). *Language socialization across cultures*: Cambridge University Press.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development, 2*(3), 230-247.

Schrödinger, E. (1944). *What is life?* Cambridge: Cambridge University Press.

Schwartenbeck, P., & Friston, K. (2016). Computational phenotyping in psychiatry: a worked example. *eneuro*, ENEURO. 0049-0016.2016.

Searle, J. (1991). Response: The background of intentionality and action. In E. L. a. R. v. Gulick (Ed.), *John Searle and his critics* (pp. 289-300). Oxford: Basil Blackwell.

Searle, J. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT press.

Searle, J. (1995). *The construction of social reality*: Simon and Schuster.

Searle, J. (2010). *Making the social world: The structure of human civilization*: Oxford University Press.

Seifert, U. (2012). Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics, 75*(12).

Seligman, R., & Kirmayer, L. J. (2008). Dissociative experience and cultural neuroscience: Narrative, metaphor and mechanism. *Culture, medicine and psychiatry, 32*(1), 31-64.

Sengupta, B., Tozzi, A., Cooray, G. K., Douglas, P. K., & Friston, K. J. (2016). Towards a neuronal gauge theory. *PLOS Biology, 14*(3).

Seth, A. K. (2014). The cybernetic brain: from interoceptive inference to sensorimotor contingencies. In T. Metzinger & J. M. Windt (Eds.), *Open MIND* (pp. 1-24). Frankfurt am Main: MIND Group.

Shagrir, O. (2006). Why we view the brain as a computer. *Synthese, 153*(3), 393-416.

Shagrir, O. (2010). Brains as analog-model computers. *Studies in History and Philosophy of Science Part A, 41*(3), 271-279.

Shapiro, L. (2010). *Embodied cognition*. New York: Routledge.

Shipp, S. (2016a). Neural elements for predictive coding. *Frontiers in Psychology, 7*.

Shipp, S. (2016b). Neural Elements for Predictive Coding. *Front Psychol, 7*, 1792. doi:10.3389/fpsyg.2016.01792

Shirkov, D. V. e., Bogoliubov, N. N., & Bogoliubov, N. (1959). *Introduction to the theory of quantized fields*: John Wiley & Sons.

Silva, P., Garganta, J., Araújo, D., Davids, K., & Aguiar, P. (2013). Shared knowledge or shared affordances? Insights from an ecological dynamics approach to team coordination in sports. *Sports Medicine, 43*(9), 765-772.

Simon, H. A. (1965). The architecture of complexity. *Proceedings of the American Philosophical Society, 106*(1965), 63-76.

Skinner, B. F. (2011). *About behaviorism*. New York: Vintage.

Solomonova, E., & Sha, X. W. (2016). Exploring the depth of dream experience. *CONSTRUCTIVIST FOUNDATIONS, 11*(2).

Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Oxford: Blackwell Publishers.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142): Harvard University Press Cambridge, MA.

Sporns, O. (2010). *Networks of the brain*. Cambridge: MA: MIT Press.

Sporns, O. (2013). Network attributes for segregation and integration in the human brain. *Current opinion in neurobiology, 23*(2), 162-171.

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B: Biological Sciences, 216*(1205), 427-459.

Stearns, S. C. (1989). The evolutionary significance of phenotypic plasticity phenotypic sources of variation among organisms can be described by developmental switches and reaction norms. *Bioscience,, 39*(7), 436-445.

Stephan, K. E., Kasper, L., Harrison, L. M., Daunizeau, J., den Ouden, H. E., Breakspear, M., & Friston, K. J. (2008). Nonlinear dynamic causal models for fMRI. *Neuroimage, 42*(2), 649-662.

Sterelny, K. (2003). *Thought in a hostile world: The evolution of human cognition.* Oxford: Blackwell.

Sterelny, K. (2007). Social intelligence, human intelligence and niche construction. . *Philosophical transactions of the Royal Society B: Biological sciences, 362*, 719-730.

Sterelny, K. (2015). Content, control and display: The natural origins of content. . *Philosophia, 43*, 549-564.

Stotz, K. (2010). Human nature and cognitive–developmental niche construction. *Phenomenology and the Cognitive Sciences, 9*(4), 483-501.

Stotz, K. (2017). Why developmental niche construction is not selective niche construction: and why it matters. *Interface focus, 7*(5), 20160157.

Stotz, K., & Griffiths, P. E. (2017). A developmental systems account of human nature. In T. Lewens & E. Hannon (Eds.), *Why We Disagree About Human Nature*. Oxford & New York: Oxford University Press.

Su, H., Wang, G., Yuan, R., Wang, J., Tang, Y., Ao, P., & Zhu, X. (2017). Decoding early myelopoiesis from dynamics of core endogenous network. *Science China Life Sciences, 60*(6), 627-646.

Sutton, J. (2007). Batting, habit and memory: The embodied mind and the nature of skill. *Sport in Society, 10*(5), 763-786.

Sutton, J. (2010). Exograms and Interdisciplinarity: History, the Extended Mind, and the Civilizing Process. In R. Menary (Ed.), *The extended mind* (pp. 189-225). Cambridge, MA: MIT Press.

Tang, Y., Yuan, R., & Ao, P. (2014). Summing over trajectories of stochastic dynamics with multiplicative noise. *The Journal of chemical physics, 141*(4), 044125.

Terrone, E., & Tagliafico, D. (2014). Normativity of the background: a contextualist account of social facts. In *Perspectives on Social Ontology and Social Cognition* (pp. 69-86): Springer.

Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*: MIT press.

Thompson, E. (2010). *Mind in life: Biology, phenomenology, and the sciences of mind*: Harvard University Press.

Thompson, E., & Stapleton, M. (2009). Making sense of sense-making: Reflections on enactive and extended mind theories. *Topoi, 28*(1), 23-30.

Thompson, E., & Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends in Cognitive Sciences, 5*(10), 418-425.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie, 20*, 410-433.

Tognoli, E., & Kelso, J. A. S. (2014). The metastable brain. *Neuron, 81*(1), 35-48.

Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, MA: Harvard University Press.

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science, 10*(1), 121-125.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*(5), 675-691.

Tomé, T. (2006). Entropy production in nonequilibrium systems described by a Fokker-Planck equation. *Brazilian journal of physics, 36*(4A), 1285-1289.

Tozzi, A., & Peters, J. F. (2018). Critique of pure free energy principle: Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Tschacher, W., & Haken, H. (2007). Intentionality in non-equilibrium systems? The functional aspects of self-organized pattern formation. *New Ideas in Psychology, 25*(1), 1-15.

Tuomela, R. (2007). *The philosophy of sociality: The shared point of view*: Oxford University Press.

Turvey, M. T. (1990). Coordination. *American psychologist, 45*(8), 938.

Turvey, M. T. (1992). Affordances and prospective control: An outline of the ontology. *Ecological Psychology, 4*, 173-187.

Turvey, M. T., Shaw, R., Reed, E., & Mace, W. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn. *Cognition, 9*(237-304).

Van de Cruys, S. (2018). Upgrading Gestalt psychology with variational neuroethology: The case of perceptual pleasures: Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychological review, 121*(4), 649-675.

Van de Cruys, S., & Wagemans, J. (2011). Putting reward in art: a tentative prediction error account of visual art. *i-Perception, 2*(9), 1035-1062.

van Dijk, L., & Rietveld, E. (2016). Foregrounding sociomaterial practice in our understanding of affordances: The Skilled Intentionality Framework. *Frontiers in Psychology, 7*(1969).

van Dijk, L., Withagen, R., & Bongers, R. M. (2015). Information without content: A Gibsonian reply to enactivists' worries. *Cognition, 134*, 210-214.

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences, 21*(5), 615-628.

Van Hemmen, J. L., & Sejnowski, T. J. (2005). *23 problems in systems neuroscience*: Oxford University Press.

Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies, 3*(4), 330-349.

Varela, F. J. (1999). *Ethical know-how: Action, wisdom, and cognition*: Stanford University Press.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*: MIT press.

Veissière, S. (2015). Varieties of Tulpa Experiences. *Somatosphere: Science, Medicine, and Anthropology*.

Veissière, S. (2016). Varieties of Tulpa experiences: the hypnotic nature of human sociality, personhood, and interphenomenality. *Hypnosis and meditation: Towards an integrative science of conscious planes*, 55-78.

Veissière, S. (2018). Cultural Markov blankets? Mind the other minds gap!: Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead et al. *Physics of life Reviews*.

Veissière, S., Constant, A., Ramstead, M., Friston, K., & Kirmayer, L. (Accepted). Thinking through other minds: A variational approach to cognition and culture. *Behavioral and Brain Sciences*.

Von Bertalanffy, L. (1950). An outline of general system theory. *British Journal for the Philosophy of science*.

Von Bertalanffy, L. (1972). The history and status of general systems theory. *Academy of management journal, 15*(4), 407-426.

Von Uexküll, T. (1987). The sign theory of Jakob von Uexküll. In *Classics of semiotics* (pp. 147-179): Springer.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological functions. In. Cambridge, MA: Harvard University Press.

Wallace, C. S., & Dowe, D. L. (1999). Minimum Message Length and Kolmogorov Complexity. *The Computer Journal, 42*(4), 270-283. doi:10.1093/comjnl/42.4.270

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological review, 20*(2), 158-177.

Weinberg, G. M. (1975). *An introduction to general systems thinking* (Vol. 304): Wiley New York.

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford: Oxford University Press.

Whitehouse, H. (2002). Modes of religiosity: Towards a cognitive explanation of the sociopolitical dynamics of religion. *Method and theory in the study of religion, 14*(3-4), 293-315.

Whitehouse, H. (2004). *Modes of religiosity: A cognitive theory of religious transmission*. Oxford: Rowman Altamira.

Williams, D. (2017). Predictive processing and the representation wars. *Minds and Machines*, 1-32.

Williamson, T. (2013). *Modal logic as metaphysics*: Oxford University Press.

Wilson, D., & Sperber, D. (2002). Relevance theory. In: Blackwell.

Wilson, D., & Sperber, D. (2012). *Meaning and relevance*: Cambridge University Press.

Wilson, R. A., & Clark, A. (2009). How to situate cognition. Letting nature take its course. In P. Robbins & M. Ayede (Eds.), *The Cambridge handbook of situated cognition* (pp. 55-77). Cambridge: Cambridge University Press.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In D. F. Jones (Ed.), *Proceedings of the Sixth International Congress of Genetics* (Vol. 1). Ithaca, New York: Brooklyn Botanical Garden.

Yoshimi, J. (2007). Supervenience, determination, and dependence. *Pacific Philosophical Quarterly, 88*(1), 114-133.

Yoshimi, J. (2011). Supervenience, dynamical systems theory, and non-reductive physicalism. *The British Journal for the Philosophy of Science, 63*(2), 373-398.

Yoshimi, J. (2012). Active internalism and open dynamical systems. *Philosophical Psychology, 25*(1), 1-24.

Yuan, R., Zhu, X., Wang, G., Li, S., & Ao, P. (2017). Cancer as robust intrinsic state shaped by evolution: a key issues review. *Reports on Progress in Physics, 80*(4), 042701.

Yufik, Y. M., & Friston, K. (2016). Life and Understanding: the origins of "understanding" in self-organizing nervous systems. *Frontiers in Systems Neuroscience, 10*.

Zahavi, D. (2003). *Husserl's phenomenology*: Stanford University Press.

Zahavi, D. (2013). Naturalized phenomenology: a desideratum or a category mistake? *Royal Institute of Philosophy Supplements, 72*, 23-42.

Zahavi, D., & Satne, G. (2015). Varieties of shared intentionality: Tomasello and classical phenomenology. In J. A. Bell, A. Cutrofello, & P. M. Livingston (Eds.), *Beyond the*

analytic-continental divide: Pluralist philosophy in the twenty-first century (pp. 305-325). New York and London: Routledge.

Zawidzki, T. (2013). *Mindshaping: A new framework for understanding human social cognition*. Cambridge, MA: MIT Press.

Zeki, S. (2005). The Ferrier Lecture 1995 behind the seen: The functional specialization of the brain in space and time. *Philos. Trans. R. Soc. Lond. B. Biol. Sci., 360*(1458), 1145–1183.

Zeki, S., & Shipp, S. (1988). The functional logic of cortical connections. *Nature, 335*, 311-317.