Anticipating Affordances

Intentionality in self-organizing brain-body-environment systems

Jelle Bruineberg

Anticipating Affordances

Intentionality in self-organizing brain-body-environment systems

Jelle Bruineberg

Anticipating Affordances

Intentionality in self-organizing brain-body-environment systems

ILLC Dissertation Series DS-2018-04

Institute for Logic, Language and Computation

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 107
1098 XG Amsterdam
phone: +31-20-525 6051
e-mail: `illc@uva.nl`
homepage: `http://www.illc.uva.nl/`

The investigations were supported by the Philosophy Research Foundation (SWON), which is subsidized by the Netherlands Organization for Scientific Research (NWO) and by a NWO VIDI-grant awarded to Erik Rietveld.

Cover photo by Martin Bruineberg.
Cover design by Nora Tinholt.
Printed and bound by Pantheon Drukkers, Velsen-Noord.

ISBN: 978-90-9030840-1

Anticipating Affordances

Intentionality in self-organizing brain-body-environment systems

Academisch Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in de Aula der Universiteit
op woensdag 7 maart 2018, te 11.00 uur

door

Jelle Pieter Bruineberg

geboren te Heemskerk

Promotiecommisie

Promotor:	Prof. dr. M.J.B. Stokhof	Universiteit van Amsterdam
Co-promotor:	Dr. D.W. Rietveld	Universiteit van Amsterdam
Co-promotor:	Dr. L. van Maanen	Universiteit van Amsterdam
Overige leden:	Prof. dr. A. Clark	University of Edinburgh, UK
	Dr. G. Colombetti	University of Exeter, UK
	Prof. dr. D.A.J.P. Denys	Universiteit van Amsterdam
	Prof. dr. M. van Lambalgen	Universiteit van Amsterdam
	Prof. dr. K.R. Ridderinkhof	Universiteit van Amsterdam
	Prof. dr. M.V.P. Slors	Radboud Universiteit, Nijmegen
	Dr. F. Russo	Universiteit van Amsterdam

Faculteit der Geesteswetenschappen

Contents

Acknowledgments

At my first day as a PhD student, a friend told me that anyone who is able to start a PhD is intellectually able to finish one, but that the secret ingredient to writing a PhD lies in "mood management". Little did I know how right she was. I am enough of a Heideggerean to know that moods are public and shared phenomena and hence that managing moods is not something you can do by yourself. I want to deeply thank all those people who, by being-there with me, co-managed *zunächst und zumeist* a positive, inspiring and stimulating atmosphere. It has been an intense and wonderful experience.

First of all, I want to thank my supervisors: Erik Rietveld, Martin Stokhof and Leendert van Maanen. Erik, you showed me, as a Bachelor student, that there is a meeting point between physics, philosophy and cognitive science. Thanks for both the guidance and freedom you offered as a supervisor. Doing a PhD with you has been a four-year adventure full of stimulating discussions, travels, philosophical road-trips and surprising materializations of our philosophical concepts. It has truly been a blast. Martin, thanks for the reassuring tone you brought to our meetings and the invaluable feedback on my writings. Leendert, you only joined the project in the last year, but your kind yet persistent insistence to find the empirical ramifications of all these philosophical constructs has been a worthwhile (and, I am afraid, still on-going) challenge. Special thanks go out to my unofficial supervisor Julian Kiverstein. Julian, you have been a great teacher and a kind critic to me both during my Masters and during the PhD. Your courses have been an indispensable stepping stone towards the success of this dissertation.

I want to thank Sanneke de Haan, Maarten van Westen, Ludger van Dijk and Daphne Brandenburg for being critics, coffee drinkers, office-mates and mood managers. The Netherlands has become a bit of a hub for young enactivists and other friendly philosophers of mind. I learned a lot from philosophizing with Katja Abramova, Carli Coenen, Roy Dings, Jasper van de Herik, Twan Kieboom, Manuel Gustavo Isaac, Mark Miller, Zuzanna Rucinska and Linde van Schuppen.

I enjoyed the rebellious atmosphere at the philosophy department of the University of Amsterdam, although teaching my first course and writing this dissertation during the Maagdenhuiszetting of 2015 was a bit intense. I want to thank (among others) Elsbeth Brouwer, Hein van den Berg, Maarten Coolen, Paul Dekker, Mathias Madsen, Sylvia Pauw, Stefan Pliquett, Federica Russo, Matth Scholten, Katrin Schulz, Christian Skirke, Alex Thinius, Hao Wang and Dilek Yamali for providing at the Oude Turfmarkt the supportive environment necessary for writing a dissertation.

Thanks to the new neighbours of the IAS, especially to Peter Sloot and Joyce the Holter, for inviting me to become an associate fellow. Having a philosophy department and a complex systems institute nextdoor from each other in the city center of Amsterdam is an incredible luxury.

Part of the research in this dissertation has been conducted abroad. I want to thank the teachers and participants of the Santa Fe Complex Systems Summer School 2015, and especially to Daniel Friedman, Tobias Morville, Matt Osmond and Susanne Pettersson for probably the most intense and exciting four weeks of my PhD. I want to thank Piet Hut for inviting me to the first ISSA summer schools in Japan and for hosting me at the IAS in Princeton, and thanks to Olaf Witkowski, Eran Agmon, Katsunori Miyahara and Elizaveta Solomonova for achieving a remarkable degree of interdisciplinarity during these events. These visits have both been humbling and rewarding.

I want to thank Anthony Chemero for hosting me at the Center for Cognition, Action & Perception in Cincinnati. I am very grateful to have been part of such a vibrant research group in ecological psychology. Thanks to Michael Richardson, Maurice Lamb, Patric Nordbeck, Ashley Walton, Hannah Douglas and Alex Jeuk for making my stay both intellectually and personally memorable.

I want to thank Karl Friston for hosting me at the Wellcome Trust Center for Neuroimaging at UCL. Immersing myself in computational neuroscience was made a lot easier by the superb guidance of Thomas Parr as well as by the long but rewarding group meetings led by Karl. Thanks as well to Micah Allen and Francesca Fardo for having me stay in their attic and for introducing me at the dinner table to their cutting-edge work in neuroscience.

The work presented in this dissertation benefited a lot from feedback and comments. I want to thank the following people for taking the time to give feedback on my work or discussing my work with me: Larissa Albantakis, Duarte Araujo, Randall Beer, Anne Brouwers, Leon de Bruin, Andy Clark, Giovanna Colombetti, Krys Dolega, Dobri Dotov, Guillaume Dumas, Regina Fabry, Sina Fazelpour, Nico Frijda, Chris Frith, Tom Froese, Shaun Gallagher, David Habets, Hermann Haken, Harry Heft, Jakob Hohwy, Dan Hutto, Hanne de Jaegher, Stefan Kiebel, Michael Kirchhoff, Michael Kimmel, Joel Krueger, Johan Kwisthout, Dimitrije Markovic, Richard Menary, Thomas Metzinger, Berk Mirza, Erik Myin, Noah Moss Brender, Alva Noe, Giovanni Pezzulo, Maxwell Ramstead, Richard Ridderinkhof, Iris van Rooij, Dan Sheppard, Peter Sloot, Mark Slors, Jonno Sutton, Wolfgang Tschacher, Nao Tsuchiya, Madeleine Thomas, Wanja Wiese, Andrew Wilson, Rob Withagen, and Dan Zahavi.

Let me also thank the journals and editors who were willing to publish some of the papers I wrote as part of this dissertation. Chapter 3 of this dissertation was previously published in the Philosophy and Predictive Processing collection by the MIND Group. I want to especially thank Thomas Metzinger and Wanja Wiese for the hard work that went into putting together the collection and I admire their goal of promoting high quality open access publishing opportunities for young scholars.

Thanks to Twan Kieboom for proof-reading the whole manuscript and and to Ronald de Haan for last minute help turning the dissertation into LaTeX. Many thanks to Nora Tinholt for designing the cover under an unreasonable time pressure, and thanks to René Jongen from Pantheon Drukkers for printing the dissertation.

Our openness to the world is shaped by a whole history of interactions, and it is therefore hard to point to one single event as having a clear causal contribution. But I doubt I would have been where I am at if not for a chance encounter at a beach bar in Wijk aan Zee in the summer of 2003. Vincent Tijms has been a role-model, mentor, friend and, eventually, colleague who showed me that the big questions of life and mind are not only subjects for late night ramblings (although still very good topics at those moments!) but can also be studied scientifically and philosophically.

Here's to a notorious bunch of mood-managers, experts at steering me away from philosophical conundrums. Dank jullie wel, Ayla, Bob, Brigitte, Chantal, Chiarli, Daan, Dave, Dries, Gerben, Guy, Jaco, Josephine, Mirthe, Roos, Sjoerd, Stannie, Stella and Willemijn. Oh, and thanks for enduring me during all those games of Risk. Dank aan mijn ouders, mijn broer Berry en mijn zus Anne. Ik ben trots op jullie.

Iris, having you come into my life halfway my PhD, instigated a series of welcome metamourphoses (not in the least the one which morphed me from couch potato to triathlete, and back...). Iris and Ronald, thanks for being such loving and wonderful people, you mean the world to me.

Amsterdam
January, 2018.

Jelle Bruineberg

Introduction

In the city center of Amsterdam, where the Staalstraat meets the Groenburgwal, stands the Staalmeestersbrug (**Figure 1**, top right). It is an old drawbridge, connecting the city hall and the flea market on the Waterlooplein with the rest of the city centre. The middle of the bridge offers an idyllic view of the Groenburgwal canal and the Zuiderkerk in the background. Since a few years, the bridge is hung with little 'love locks', sold by a clever shop owner at the corner. The bridge facilitates many activities: attaching a lock to it and throwing the key in the water, taking a picture, selfie or even a family portrait, crossing by bike or by foot or, mainly accidentally, by car. On a typical sunny afternoon, these activities take place simultaneously on the surface of the bridge.

I have crossed this bridge many times during the daily commute to my office at the University of Amsterdam. Over the years, it has become a bit of a hobby of mine to observe and participate in the meeting of activities taking place at the bridge. The romantic couple, the eager photographer, and the impatient cyclist each have different concerns, a different pace and a different focus, each only in part attentive of the other (**Figure 1**, top left). For example, the worlds of the cyclist and the photographer are moving literally orthogonal to each other: the cyclist aligns with the road and crosses the bridge, the photographer aligns with the canal and crosses the road, forming a meshwork at the bridge (**Figure 1**, bottom right).

One can often observe an all too familiar scene taking place. One family member, usually the father, arranges the children and spouse and, keeping the family firmly in view through the display of his photo camera, walks backward searching for the perfect constellation of family, bridge, canal and church. The cyclist, on collision course with the oblivious photographer, rings her bell mildly agitated and manoeuvres her bike in between the frightened photographer and the stunned family. While biking away the cyclist can't help but grin at the number of photos she has bombed in this way.

Figure 1: The Staalmeestersbrug and its affordances. Top left: Pedestrians and cyclists meeting at the bridge. The cyclist on the left almost runs into the posing couple. Top right: The bridge and the views it offers. Bottom left: municipality workers removing the locks with a grinder. Bottom right: A cyclist manoeuvring in between a photographer and the one posing for the photo.

Situations like these are typical of everyday human activity. There are many different ways to analyze such activities, each highlighting a different aspect of the situation. In this introduction, I will articulate four different aspects, each will be further developed in the chapters to come: *socio-material*, *normative*, *dynamical* and *phenomenological*.

The *socio-material*[1]perspective includes an understanding of the different action possibilities, or affordances, offered by particular places: the bridge affords crossing by foot and by bike, taking a photograph and attaching a 'love lock', and many more activities. In turn the different affordances of the bridge are made possible by both the physical layout and situatedness of the bridge, and the practices the different actors are part of. If the bridge would not offer such a nice view, connect different parts of the city, or have steel cables to which locks can be attached, it would not have the affordances that it has. On the other hand, if people would not have eyesight, ride bicycles, have photo cameras and know how to use a camera, the bridge also would not have the affordances it has. Affordances are therefore founded upon the layout of the environment, the abilities available to an agent and artefacts and technologies.[2] Since all these factors are to a greater or lesser extent in flux, affordances themselves are changing and open-ended. Importantly, although the affordances the environment offers are rich, abundant and open-ended (think of what the bridge offers to a parkour runner, an urban climber and a modern dancer), they are also limited: the bridge does not afford picking up, drinking from, or marrying.

In a different *normative* way, social practices constrain affordances. Some are loosely constraining: the practice of making holiday pictures involves both the family and what-there-is-to-see to be in the same picture. Other practices constrain in a more specific, or ritualistic, manner: one only 'validates' one's relationship if one writes the right initials on a lock (rather than someone else's), if one attaches it to the bridge (rather than one's backpack) and throws away the key (rather than stick it in one's pocket). Human activities like these unfold within a practice in which it is appropriate or inappropriate to do something, i.e., they are inherently normative. We often correct and guide someone else's behavior: we say "look into the camera", "you are supposed to throw the key in the water", or ring our bell on the bike when someone stands in the way.

---

[1] With the term *socio-material*, I want to point at the intertwinement of the *material* layout of the environment with *socio-cultural* practices. This can be understood in the rather trivial sense that the material environment surrounding us is a product of (human) activities. I want to go for the slightly stronger claim that the affordances which an aspect of the material environment offers to a human can not be understood apart from their situatedness in socio-cultural practices. I want to remain neutral, and I think I can, about the stronger claim that materiality itself is socially constituted (see van Dijk & Rietveld, 2017; Mol, 2002; Orlikowski, 2007).

[2] More precisely, as we will see in Chapter 2, affordances can be defined as relations between aspects of the (socio-material) environment and abilities available in a form of life (Rietveld & Kiverstein, 2014).

Practices themselves are *dynamic* and can be contested as well. The love-lock practice spread from the Pont des Arts in Paris to many bridges and places all over the world. It is an open question what facilitates the success of such practices and makes them stick at some places and not at others. Elements of an answer might be the concrete layout of the bridge, the romantic atmosphere of the canals of Amsterdam and the shop at the corner selling the locks. Some inhabitants of the neighborhood have complained to the municipality about this practice and have warned that the bridge might not be able to support the load. This has resulted in workers of the municipality removing the locks every few months (**Figure 1**, bottom left). One of the latest developments on this matter is a little article in a Dutch newspaper (Kruyswijk, 2017) stating that the municipal councilor of infrastructure is negotiating with Google to remove the name “love-lock bridge” from Google Maps. At the time of writing of this introduction, November 2017, the name of the bridge in Google Maps has in fact been changed to “staalmeesterbrug [sic]”.

One can analyze the dynamics at the bridge at multiple time-scales: the emergence, decline and blending of practices (100 years ago lovers did not attach love-locks to bridges, or, at least, did not take selfies with them), the building of the bridge and the continuous wear and repairs, the unfolding of a concrete situation etc. These time-scales are interrelated: because of the emerging practice of hanging love-locks at the bridge, the locks afford removing for the government worker on Monday morning, which in turn preserves the bridge for the future. As such, one might speak of a “stability in flux”, a more or less stable socio-material practice at a particular place, enabled, maintained, threatened and supported by a multiplicity of activities.

All these aspects are at play in a concrete situation. When arriving at the bridge while a photo is being taken, the cyclist has a number of alternative options available. Some might start ringing their bell from afar, sometimes accompanied by a shout, flagging the danger and signaling the photographer they are in their way. This typically leaves enough time for the photographer to move to the side. Others only ring their bell at the last moment, refusing to slow down, leading to the photographer jumping away (or in the worst case in front) of the bike. One proven tactic is, if there *is* an opportunity to pass, to not ring at all in order to prevent people from making sudden movements. Which options should be taken might depend on subtle changes in context; navigating one’s way in between the photographer and the posing family is only possible when arriving at a particular point in the photo-taking dynamic, showing the intertwinement of the *socio-material*, *normative* and *dynamic* aspects of everyday activities.

The situation shows up in specific ways in the *phenomenology*[3] of an agent at home in the situation. The photographer stepping back on the road is drawn towards an optimal constellation of the bridge, the family and the church in the background. The cyclist is drawn towards the looming gap between the photographer and the posing family. While the experienced cyclist anticipates the unfolding of the situation, passes the family and continues on her way, the photographer might be struck by the for him unexpected passing of the cyclist, realizing only then that he is standing in the middle of a bicycle path. Both the cyclist and the photographer attune to the situation in a specific and different way. Each has a distinct state of action-readiness specific to the concrete situation and a corresponding *field of relevant affordances.*[4]

A central idea that will be developed in this thesis is that through a history of interactions, agents develop a sensitivity for the relevant action possibilities in a situation. The relevant action possibilities, or solicitations, show up in their phenomenology as tensions that need to be reduced. By being ready for and acting on the relevant action possibilities, an agent tends towards grip on the environment. Due to the encounter with the cyclist, the field of affordances of the photographer is now changed. It might now involve an openness to approaching cyclists as well, perhaps leading him to pause for a moment until the next cyclist has passed. An agent's history of interactions with the environment shapes the agent's openness to affordances. If these changes persist over time, we can say the agent has now learned something about or is better attuned to its environment. I will call an agent's directedness at its environment *skilled* [5] *intentionality* and the actions manifested in this way *skilled action.*

The above bridge example shows the extreme richness of skilled action, the main topic of this thesis. It is also supposed to make plausible that a single *explanans* will be highly unlikely: a satisfactory explanation of skilled action will, at the very least, have to appeal to practices and norms, the regularities in the environment, an agent's history of interactions with the environment, an agent's phenomenology and the internal (neuro)-dynamics of the agent. The aim of this thesis is therefore to develop a *conceptual framework*, that shows how these different aspects fit together, and to develop a set of concepts that can be put to use in the various fields that relate to each of these different aspects of skilled intentionality.

---

[3] I use the term *phenomenology* in a colloquial sense here, pertaining to the structure of the experience of an agent.

[4] Varela (1999) calls such states of action-readiness a micro-identity and the resulting lived situation a microworld.

[5] As will become clear throughout this introduction, the term *skill* in *skilled intentionality* does not demarcate one form of (unreflective) intentionality from other (more reflective) forms of intentionality, but rather characterizes all forms of intentionality performed by skilled agents.

Outline of the introduction

This thesis contains two complementary lines of argument. On the one hand, it is an attempt to develop a conceptual framework for the study of skilled intentionality in terms of dynamical systems theory and predictive processing. On the other hand, it presents an argument that predictive processing, and especially its systems theoretic cousin, the free-energy principle, are best understood in terms of an ecological and enactive philosophy of mind, and not in the rationalist philosophy of mind it is standardly associated with. The first line builds on and enages with previous work in radical embodied cognitive science (Chemero, 2009), autopoietic enactivism (Varela, Thompson, & Rosch, 1991; Thompson, 2007), Heideggerian cognitive science (Dreyfus, 2007; Freeman, 2000; Kiverstein & Wheeler, 2012) and radical enactivism (Hutto & Myin, 2013, 2017). The second line builds on and engages with previous work on the free-energy principle (Friston & Stephan, 2007; Friston, 2013b) and the philosophy of predictive processing (Clark, 2013, 2016b; Hohwy, 2013; Metzinger & Wiese, 2017).

One key thing I learned while working on this dissertation, is that the appraisal of the plausibility of a framework is heavily dependent on one's philosophical intuitions. Those who share my embodied and ecological philosophical intuitions might see the endeavor of this thesis as the liberation of a promising theoretical principle from the hands of a deeply problematic view of the mind. Those who do not think this view of the mind is problematic in the first place, might see a contrived and convoluted attempt to twist the predictive processing story in such a way as to align it with (radical) embodied cognition.[6] The consequent debates often focus on the use of specific concepts used in predictive processing such as "Markov blankets", "inference" and "prediction-error" and are more often than not inconclusive. What plays out on the background, and rarely becomes explicit, is a clash between two fundamentally incompatible ways of understanding the mind-world relationship that are frontloaded in each of the different interpretations. I think that frontloading this clash will be helpful to make intelligible the arguments that follow.

I will therefore continue this introduction by bringing to the fore this dominant understanding of the mind-world relationship, show its main problems, and lay out the possibility for an alternative non-intellectualistic understanding of the mind-world relationship. I will then present how Freeman's work on neurodynamics can support this non-intellectualistic understanding of the mind-world relationship and show the continuity between Freeman's work and Friston's work. I then introduce the philosophical strands most closely neighboring on the current research project: *Radical Embodied Cognitive Science*, *Radical Enactivism*, and *Autopoietic Enactivism* and briefly point to how they relate to the proposal put

---

[6] Or, as Jakob Hohwy, one of the main antagonists of this thesis, confided to me after a few drinks: "I really don't get why anyone would still talk about affordances".

forward in this dissertation. In the last part of this introduction, I will address some of the methodological issues that come along with the articulation of an interdisciplinary framework.

Setting the philosophical stage

In their recent book *Retrieving Realism*, Dreyfus and Taylor (2015) give a characterization of a particular picture of the mind and the mind-world relationship that, they claim, pervades all through the philosophy of mind and cognitive sciences.[7] This picture, they argue, has its roots in the work of Descartes and Locke, and although virtually nobody today subscribes to Descartes' substance dualism or Locke's concept empiricism, it is "a picture [that] held us captive" (Wittgenstein, 1953, §48). That is to say, it is a picture that, in a sense, feels so natural that it is hard to think outside of it, but still imports a specific, and in the view of Dreyfus and Taylor highly problematic, topology of mind and world. The logic of this topology is simple: "[t]he reality I want to know is outside the mind; my knowledge of it is within. This knowledge consists in states of mind which purport to represent accurately what is out there. When they do correctly and reliably represent this reality, then there is knowledge" (Dreyfus & Taylor, 2015, p. 2). The particular dualist sorting of mind and world that Descartes laid out in his *Meditations on First Philosophy* (Descartes, 1641/1988) has been refuted over and over again, but the more provocative aspect of Dreyfus and Taylor's analysis is their claim that despite this refutation, the deeper topology dominates philosophical and (neuro-)scientific thinking up until this day.

This topology is characterized by four interwoven strands. The first strand is what Dreyfus and Taylor identify as the "only through" structure: we have access to the "outside" world *only through* the bounds of the mind/organism. This bound then effectively screens us off from the world, our knowledge of the external world is mediated by whatever happens at these boundaries. From this it follows that in order to have knowledge of the world at all, we, or our brains, need to infer, based upon these mediational states, what the state of affairs of the world really is.

Second, our knowledge consists of separate and clearly defined entities, whether these are 'clear and distinct' ideas, beliefs, or sentences held true. What is important here is that in all cases knowledge is understood as articulated, explicit and accessible from the first person. Knowledge is therefore seen as a form of knowing-that.

---

[7] The standard picture of the mind is described both by its proponents and opponents. I choose to focus on Dreyfus and Taylor's characterization of it since it is unusually clear and, I think, presented neutrally.

Third, in justifying our beliefs, an agent is bounded by and can not reach beyond the bounds of the mind/organism, it can only make reference to what impinges on it.

The fourth strand Dreyfus and Taylor identify is the dualist sorting between the mental and the physical. This is not to be understood as an ontological claim – substance dualists are rare these days – but rather as a conceptual one. The materialist "accepts the sorting, but claims that only one term is really instantiated" (p.11). The basic Cartesian sorting between a material realm governed by substance and causality, and a mental realm governed by beliefs and rationality remains intact. The aim of the materialist program is to give the mental a material explanation, but the mental here is conceived of as an *inner realm* made up of distinct entities or ideas, hence leaving the original sorting intact.

These four strands rehearse in an elegant way Dreyfus' long standing critique of the rationalistic and mentalistic foundations of artificial intelligence (Dreyfus, 1992/1972) and it is an open philosophical question to what extent modern epistemology and philosophy of mind are still captives of this picture. The point of presenting these ideas here in the introduction of the thesis is, first and foremost, to show that the mediational view is a substantive hypothesis rather than a logically necessary starting point.

There is a sense in which it is trivial that our brains are not directly connected to the rest of the world and that they are spatially insulated by our skull. However, it does not follow from this that therefore our skulls "screen off" our minds from the world and we therefore can gain knowledge of the world only indirectly (for one, it would follow that our minds can be put in more direct contact with the world through trepanation). It is only when one assumes that we are locked 'inside our head' and that we need to reach out to rest of the world to obtain knowledge, that spatial seclusion becomes epistemic seclusion. Furthermore, as humans we know a great number of *facts* about the world, but from this it by no means follows that all of our knowledge of the world has a fact-like representational structure. It is this self-perpetuating basic picture that Dreyfus and Taylor try to make us aware of. Peeking ahead, when it comes to the philosophy of predictive processing in particular I think it is extremely tempting, but ultimately wrong, to apply this topology to the models that are proposed.

Agency, cognition and purpose

The topology that holds for perception - there is a world out there and our knowledge of it is within - holds for action as well. Our reasoning about what to do, fueled by our beliefs about the world and our desires, happens inside, while the purported consequences of our actions are out there in the world. We do not need to invoke the pineal gland anymore to connect the mental with the physical (the brain took the place of *res cogitans*), but much of the logic remains the same.

On the standard account, the difference between acts, performed by agents, and mere mechanistic changes in the world is that acts are events that stand in the right causal relation with mental states (Davidson, 1963).

Descartes famously thought that animals lacked such a capacity and could be understood completely in mechanistic terms. He held that particular human behaviors, such as reflexes and the early stages of perception, could be understood in purely mechanistic terms as well (Descartes, 1633/2011). But more important than the particular locus of the borderline is the dichotomy between intentional agency and blind mechanism itself, together with the claim that nature can be exhaustively described by a combination of both. In other words, behaviors come in only two flavors: they are *either* mechanistic and devoid of purpose produced by mere automata, *or* cognitive, intentional and goal-directed produced by full blown cognitive agents (see Withagen (2013) and Fulda (2017) for further discussion).

Descartes held that genuine cognition and mere mechanism can be dissociated by the flexibility of behavior, for example the ability to produce sensible streams of words as a response to a wide range of things that are being said. Furthermore, even though machines or animals might match or exceed human ability in some respect, they would fail in others. Reason is flexible and domain general, whereas reflexes are static and domain specific (Descartes, 1637/1979). Hence, we can infer cognition based on the presence of an organism's flexible or domain general abilities.

The equation of intentional behavior with representation and cognition is not an eccentric 17th century position, but rather the mainstream view to date. For example, Dretske writes:

> To describe the hen, for example, as engaging in diversionary tactics (to protect her chicks) is already to describe her behavior in a way that presupposes an intentional structure for the internal source of that behavior. The appropriateness of response, then, insofar, as this is relevant to what the organism believes and intends, is a property the response acquires only in virtue of its production by internal states having content. (Dretske, 1980, p. 284)

Here we see the corollaries of the mediational view at work: purposive and flexible behaviors presuppose an internal source, namely contentful states such as beliefs, desires and intentions. As said, it is not where Descartes draws the distinction between mechanism and reason, but rather the exhaustiveness of the distinction that has been most influential.[8]

---

[8] In this sense Cartesian dualism, classical computationalism (Putnam, 1962) and anomalous monism (Davidson, 1970/1980) are all very different ways of fitting the causal and the mental realm. What they have in common is that the causal and the mental (understood as internal states having content) are the only two pieces of the puzzle.

Accepting that these are the only two options leads to theoretical tensions. The observation that there is adaptive and, seemingly, intelligent behavior forces one into a particular kind of characterization of the internal goings-on of a system. For example, Fulda (2017) presents some of the terminology used by leading researchers in the field of bacterial cognition. Baker and Stock (2007) conclude a short position paper on networks and integrated circuits in bacterial cognition by stating that: "[t]he next task is to understand how [genes and proteins] are connected to form a dynamic, adaptive cell" (2007, p. 1023). While this is a perfectly legitimate scientific question, they spell out this task in the following way: "How is information converted into knowledge, and how is knowledge sorted, evaluated and combined to guide action, morphogenesis and growth?" (2007, p. 1023). The implicit assumption here is that adaptive behavior and knowing-that go hand in hand, where one sees the former, the latter is implied.

It is unclear whether these researchers want to claim that bacteria *literally* deliberate about what to do based on the information at their cell wall or whether they merely see this as a useful *metaphor*. But the question is how well a framework designed for detached human reasoning processes is suitable to explain something as basic as bacterial chemotaxis and whether, in making the framework fit, its original terms have not come to have a completely different meaning. Using the same terms for different phenomena is in itself scientifically problematic, but when one usage of these terms imports a highly specific topology of mind and world, one should be extra careful.

Towards a middle ground

By accepting the dichotomy between causal and intellectualistic language, one is forced to either reduce or intellectualize agency. In doing so, one excludes the possibility to discover and conceptualize forms of agency and intentionality that do not meet the intellectualistic criteria, but that are nevertheless adaptive and, perhaps, experiential. This dichotomy between causes and reasons is judged to be problematic by philosophers from both the analytic and continental tradition, both when it comes to non-human agency and when it comes to skilled human agency. For example, Davidson in his article *The Emergence of Thought* writes:

> We have many vocabularies for describing nature when we regard it as mindless, and we have a mentalistic vocabulary for describing thought and intentional action: what we lack is a way of describing what is in between. This is particularly evident when we speak of the 'intentions' and 'desires' of simple animals; we have no better way to explain what they do. (Davidson, 1999, p. 11)

Davidson resists attributing judgements and beliefs to non-human animals, for he thinks that such capacities require the mastery of linguistic skills. In continental philosophy, Merleau-Ponty touches upon the issue in the context of articulating the findings of Gestalt psychology. According to Merleau-Ponty, by being committed to the dichotomy between reasons and causes, the Gestalt psychologist cannot express its own findings, the phenomenon of a non-explicit, not fully determined (in Merleau-Ponty's words *non-thetic)* perception;

> Objective thought cannot assimilate these phenomena, and this is why Gestalt theory (which, like every psychology, is a prisoner of the "facts" of science and of the world) can only choose between reason and cause, and why every critique of intellectualism ends up (in the hands of objective thought) in a restoration of realism and of causal thinking. On the contrary, the phenomenological notion of *motivation* is one of those "fluid" concepts that must be formulated if we want to return to the phenomena. (Merleau-Ponty, 1945/1962, p. 51)

Merleau-Ponty uses the concept of a *motive* as a relation that is implied by a particular organization. One particular class of motives are motives to act: when a stranger stands too close in an elevator, one experiences a tension that stands in need of being reduced, the situation *motivates* an action, stepping away, that results in a new, temporarily stable equilibrium. In this thesis, I will follow Merleau-Ponty and thinkers like Hubert Dreyfus, Alva Noe and Sean Kelly in conceiving of *motivations* or *solicitations* as involving an intentional relation that is *sui generis*, prior to and distinct from the traditional characterization of intentionality as desire or belief, and able to capture non-human agency as well as central cases of human agency.

As will be articulated in more detail in *Chapter 1* of this thesis, the basic intentional picture is that for a particular system with a set of abilities *s*, in a particular situation or context *c*, a number of aspects *a* of the environment - or affordances - solicit or motivate an action. In a different context *c',* or for a system *s'* with a different organization and different abilities, different aspects of the environment *a* solicit action. The context-sensitivity and skill-dependence of motivations is what distinguishes them from mere reflexes. The second component of the basic intentional picture is that an agent's openness to the world is characterized by what Merleau-Ponty calls *the tendency towards an optimal grip.* An agent well adapted to its niche will be solicited by those affordances that, when responded to, lead toward an improved grip on its environment.

In this thesis, I call the tendency towards an optimal grip a "basic concern of living organisms". This is not just a brute ontological posit, but rather the outcome of an analysis of what it is to be a living system. The very fact that living systems resist a natural tendency to disorder puts specific constraints on their interactions with the environment. This starting point is shared by a disparate set of thinkers discussed in this thesis - ranging from Hans Jonas to Karl Friston. This sensitivity is partially structured as resulting from extrinsic forces: organisms without the proper relation to their environment will simply have died off (Maturana & Varela, 1987). But in order to ensure *adaptive* actions, that is to say, to ensure that the action possibilities that solicit will lead to an improved condition for the organism, the agent needs to be sensitive to its own conditions of viability and flourishing in regulating its interactions with the environment. This results in a kind of teleology specific to living systems, whether understood as self-concern, the tendency towards an optimal grip, or the minimization of free-energy.

The picture that emerges then is that of a living system tending towards grip on a situation by being selectively open and responsive to a multiplicity of affordances, i.e., by skilled intentionality. In this thesis, I will develop a framework for the naturalization of skilled intentionality. By naturalization here I don't mean a reductive explanation, but rather an attempt to kickstart a productive cognitive science based upon the principles and concepts used in this dissertation. If this is possible I will kill two birds with one stone. On the one hand, the resulting cognitive science will avoid the pitfalls of the mediational picture that Dreyfus and Taylor point out (it is therefore of utmost importance not to reintroduce them in explaining skilled intentionality). On the other hand, we will have found a stable middle ground between mechanism and intellectualism, providing the conceptual tools for an explanatory account of agency and intentionality that does not presuppose representation and deliberation.

Intellectualism and the intellect

I have argued that an account of intentionality modeled after our intellectual engagement with the world is ill-suited to capture skilled human agency and non-human agency. Of course, it is then an important question whether and how skilled intentionality is able to capture our more intellectual engagements with the environment. This question is at the very heart of various strands of embodied cognitive science, be it guised as the problem of "representation-hungry cases" (Clark & Toribio, 1994), the "intelligible interface problem" between basic minds and contentful minds (Hutto & Myin, 2013). I will not attempt to solve this issue here, nor anywhere else in this thesis (I think it would require, at least, a dissertation of its own), but I think that, at least, a positive outlook can be given.

One influential, and ultimately problematic, starting point for this discussion is Heidegger's distinction between the *ready-to-hand* and the *present-at-hand.* For Heidegger (1996/1927), there are two fundamentally distinct modes of access to the world (or modes of being). The *ready-to-hand* is associated with skilled, fluent and immersed engagement with the environment, the *present-at-hand* is associated with a detached, objectifying and scientific observation. Heidegger's own timeworn example concerns the case of a hammering. When we are hammering away skillfully the world and the hammer "withdraws" or "disappears" from our experience, it does not show up at all. It is only when things go wrong, the head of the hammer flies off, that the world becomes present and explicit.

This distinction is often used, as in Dreyfus' influential Heidegger interpretation (1991), to argue for the priority of immersed activity over detached thought: "all thematic [conceptual] intentionality must take place on a background of transparent coping. In order even to act deliberately we must orient ourselves in a familiar world" (Dreyfus, 1991, p. 85). I think that this way of putting the distinction is unhelpful in a number of ways. It leaves our skilled interactions devoid of any mindedness and over-intellectualizes our more intellectual engagement with the world. As Alva Noe aptly notes:

> Such a conception of the intellect - as modality of detachment rather than a modality of openness to the world, to use McDowell's (1994) phrase - violently distorts our lives as thinkers, perceivers, and doers. [...] Heidegger, and Dreyfus, repudiate presence in favor of absence, because they insist that there can be no "unthought" presence, and they insist on this because they take for granted an over-intellectualized conception of the intellect, just that conception that modern philosophy has taken for granted. (Noe, 2012, p. 9)

Rather, Noe holds, "the whole field of the intellect - thought, concept-use, language - is itself a sphere of absorbed coping (to use Dreyfus' phrase)" (2012, p. 9). Our openness to the world - what Noe calls presence - does not come for free. It is made possible by "a very complicated, hard-won set of skills" (p.10). It is the differences in these skills that make the difference between hearing a sound, hearing a string of words, hearing a sentence in Spanish and hearing the opening line of *Cien anos de soledad.* Although the sound wave might be the same, in each case we are presented with a different experience.

At the same time, in episodes of unreflective action, the world does not completely disappear from our phenomenology. It is exactly in these episodes that we are sensitive to the demands of the situation. Our openness to the world is structured by the norms that hold in the practice we grow up in and by our history of interactions with the environment, which shapes our abilities and what we care about. The skilled Amsterdam-style cyclist from the opening paragraphs navigates the bridge with an openness that might be characterized as being assertive (if not aggressive) but all the while being sensitive to the photographer's inexperience with the situation. At the same time the bike, including the rattling sounds and squeaks, is pushed to the background of the cyclist's phenomenology.

What lays before us then, at least as a viable working hypothesis, is the idea that skilled intentionality, understood as selective openness to multiple affordances simultaneously, can capture all human and non-human agency. What differs are the skills possessed by the agent and the practices the agent partakes in. Higher cognition and our intellectual engagement with the world can all be understood as engagement with *just more affordances.*[9] The account of *general*

---

[9] Indeed, as I will discuss in Chapter 4, this characterization puts pressure on the very distinction between 'lower' and 'higher' cognition.

ecological information, as developed in *Chapter 4*, is designed specifically to do this work. However, justifying this claim is something that would require a number of worked out examples of how cases of higher cognition, such as design, language use and writing a text can all be understood as cases of selective engagement with affordances. That work is not part of this thesis, but is currently carried out by other members of the VIDI/ERC Research Group on Affordances for 'Higher' Cognition & Skilled Intentionality (see for a start Rietveld & Kiverstein, 2014; Rietveld & Brouwers, 2017; van Dijk & Rietveld, 2017; Van Dijk and Rietveld, under review; Van Westen, Rietveld and Denys, under review; Kiverstein and Rietveld, in preparation).

Relevance, goals and frames

The basic starting point of this dissertation is that the environment is *replete* with affordances. The main problem for the agent is therefore to be *selectively* open to affordances in ways that are beneficial for the agent, it needs to be open to the *relevant* affordances. Much is packed into the notion of relevance: what is relevant depends on the general setting, the practices an agent partakes in, specific details of the situation. Even more importantly: what is relevant at one moment might not be at the next: when the biker whizzes by, the photographer is no longer drawn by the optimal constellation of the photograph, but rather drawn to the safety of himself and his family.

Again, there is a clear intellectualistic framework available to conceptualize relevance: what is relevant depends on our *goals* and *intentions*. The photographer has a number of goals in mind, making a good picture and keeping the family save might be two of them. These goals might be ranked according to a hierarchy (Maslow, 1943) with safety being more basic than holiday pictures. Relevance is then determined by some sort of counterfactual reasoning given a particular goal: "what actions are to be done if I want to make a good picture", until some sensory input indicates that a more basic goal, "safety" is threatened, which then comes to guide behavior.

The main challenge for such an approach is that it is unclear how and why some goals are prioritized in a situation-specific manner. We need to explain why we have the goals we have and why some of them are ranked higher than others and why we change the prioritization when something in the environment changes. The bet of *Heideggerian* and *Radical Embodied Cognitive Science* is that an approach that starts from explicit facts, contexts and goals will eventually run into the frame problem (Dreyfus, 1991; Wheeler, 2008; Rietveld, 2012b; Shanahan, 2016): in order to disambiguate which context applies the agent needs to rely upon environmental cues. But knowing what cues of the environment to rely upon already presupposes that one understands the context. Understanding that the approaching cyclist is the relevant piece of information rather than the railing of the bridge, or the person looking down at the situation from their balcony, is already to have solved the hardest part of the problem.

According to Dreyfus, the frame problem is a problem specifically for accounts of cognition that start from discrete sensory elements and discrete goals and from there start to compute what piece of information is relevant in which context. In other words, it is a problem specifically for the *mediational* worldview. The reason why human agents don't run into the frame problem is because they "realize the property of thrownness" (Wheeler, 2008, p. 1). Even without the Heideggerian baggage, the pressing question is how "our present concerns and past know-how always already determine[...] what will be ignored, what will remain on the outer horizon of experience as possibly relevant, and that will be immediately taken into account as essential" (Dreyfus, 1991, p. 263). The question is what kind of internal organization of (human) agents makes such relevance-sensitivity possible.

Neurodynamics and anticipation

Drawing upon work by neuroscientist Walter Freeman (1987, 2000), Dreyfus (2007) proposes a holistic and integrated purposive organization that constitutes our openness to affordances. Based upon years of work on the workings of the olfactory system in alert and moving rabbits, Freeman (2000) developed a model of rabbit learning through a history of interactions with the environment. He understands the brain to be a non-linear dynamical system: dependent on the current state of activiation of the cortex, the same stimulus can die out or greatly perturb the dynamics of the system. The same odor of a carrot can, dependent on the current activation pattern of the cortex have radically different effects. The state space of the cortex can be regarded as an attractor landscape with multiple basins of attraction, each directly related to an action. What is crucial here is that "each new attractor does not *represent*, say, a carrot, or the smell of carrot, or even what to do with the carrot. Rather, the brain's current state [...] is directly *coupled with* or *resonates to* the affordance offered by the current carrot" (Dreyfus, 2007, p. 258, last two italics mine). Dreyfus calls this "direct perception of significance" (p.259).

The integrated purposive organization takes the shape of an attractor landscape that captures the significance, or relevance, of affordances. The same patterns that embody the significance of affordances directly "guide the motor systems into sequential movements of intentional behavior" (Freeman, 2000, p. 114). An animal (in a particular macroscopic pattern) perceiving a significant stimulus is directly "readied" to act on relevant affordances.

On this model, learning to respond to a new odor is not a matter of adding an extra 'fact' to a repository, or to learn a stimulus-response pairing. Rather, "[w]hen an animal learns to respond to a new odor, there is a shift in *all* other patterns, even if they are not directly involved with the learning" (Freeman, 2000, p. 22, my italics). In other words, learning a new behavior involves a *reorganization* of the whole attractor landscape. It is because of this holistic organization that relevance and context are already intrinsic to the system from the very beginning.

The basic picture Freeman provides is that of a stimulus *perturbing* the action-readiness pattern of the brain. Through repetition this might lead to a new organization of the system that changes the way in which the same stimulus perturbs the system. This account can easily be contrasted with an account that takes the processing of a stimulus into a representation as its starting point: while in the latter case, the stimulus is *transformed* into a representation, in Freeman's case the stimulus *perturbs* the intrinsic dynamics of the brain.

Context and switching

Based on the previous analysis of skilled intentionality, one can add extra constraints on the dynamics of the brain. It needs to sensitive to the context and history: a bottle of water affords drinking when I just put it there and I am thirsty, but not when my colleague just put it there. There is no need to remember the 'fact' that it is my colleague's bottle, but it *does* need to shape my openness to affordances.

A second, and related, aspect of skilled intentionality is our openness to multiple affordances. According to Dreyfus, the horizon of experience is structured according to: "what will be ignored, what will remain on the outer horizon of experience as possibly relevant, and that will be immediately taken into account as essential" (Dreyfus, 1991, p. 263). When typing these sentences, I am solicited by the keyboard and the screen, slightly more in the background are a bottle of water, my smartphone and a pile of philosophy books. But there are also happenings on the horizon of my field that are not sensorily present: a colleague that might come in, a phone call I might get. When these events occur, my field of affordances restructures itself and attunes to the new relevant aspects of the situation: asking my colleague how she's doing or picking up the phone.

From a neurodynamic perspective, this flexible *switching* in action-readiness patterns is made possible by *metastability* (Kelso, 1995; Friston, 1997). As I will describe in *Chapter 1*, metastability can be understood as the outcome of two competing tendencies: the tendency of the components to couple together and the tendency to express their independent behavior. In a metastable regime, the system is poised at the edge of instability, a kind of dynamic stability that allows the system to maintain "a balance in the readiness of the system to transit between multiple attractors" (Davids, Araújo, Hristovski, Passos, & Chow, 2012, p. 119).

Anticipation and improvement

Flexibly switching between action-readiness patterns is one important ingredient for skilled action, but switching is in itself not necessarily adaptive. The agent is required to be sensitive to how well it's faring in its interactions with the environment, and switch accordingly. Dreyfus writes:

> The animal must take account of *how things are going* and either continue on a promising path, or, if the overall action is not going *as well as anticipated*, the brain self-organizes so the attractor system jumps to another attractor (Dreyfus, 2007, p. 259, my italics).

On this model, whether to switch behaviors or not (and where to switch to) is dependent upon how well an agent takes itself to be faring in its interactions with the environment, and how well it is faring is understood relative to what the agent *anticipates*. Without a proper account of *anticipation*, it remains mysterious how an agent can come to tend towards grip. It is not *prima facie* clear how an agent can come to anticipate the consequence of its action and switch to a new action if the action is not going as well as anticipated. Surely an agent can generate anticipations and change its behavior, but how does it come to generate the *right* anticipations and change to the *right* behavior?

One needs to be careful here. It is fine for Dreyfus and Freeman to claim that their account of brain dynamics is aimed to be "structurally isomorphic" (Dreyfus, 2007, p. 259) with Merleau-Ponty's notion of the *tendency towards an optimal grip*, but it would be a *very* costly move to suggest that the phenomenological concept of optimal grip *explains why* we act in the way we act.[10] Another costly move would be to attribute optimality conditions to self-organizing system in general (see Schneider and Kay (1994) for a strong account of "teleological physics" and Haken and Tschacher (2010) on why this might not be the best characterization of self-organization). Exactly how it comes that neurodynamics self-organizes so as to make an agent tend towards grip on its environment is a pressing question.

Self-organization and the labile brain

One of the greatest advantages of Freeman's work is that it is built around and exploits the fact that the brain's dynamics are intrinsically instable and form transient patterns over multiple spatial and temporal scales. Self-organization and the intrinsic instability of the brain's dynamics are an increasingly important

---

[10] With "costly" I mean here that the 'assumption' that a phenomenological organization is able to support or maintain its isomorphism with a neural organization is either inconsistent or a very dubious metaphysical posit. See Section "On the role of assumptions" in conceptual frameworks.

object of study in neuroscience (see for example Rabinovich, Friston, & Varona, 2012; von der Malsburg, Phillips, & Singer, 2010). In an article co-authored by many pioneers in the field, self-organization is characterized as:

> [T]he spontaneous formation of patterns and pattern change in systems that are open to exchanges of information with the environment and whose elements adapt to the very patterns of behavior they create. (Engel et al., 2010, p. 268)

The two leading questions are then how to best *characterize* the dynamics of the brain, and how to think about the *functionality* of the dynamics of the brain in connection to the interactions an agent has with its environment.

Characterization: a hierarchy of time-scales

According to (a relatively early) Friston (2000a), the dynamics of the brain can be characterized from two complementary perspectives, which he calls *type I* and *type II* complexity. Type I complexity characterizes brain dynamics as one dynamical system with a complex attractor manifold. This dynamic might have qualitatively different characteristics, dependent on where on the attractor manifold the system currently is, but these qualitative changes are only *apparent.* Type II complexity characterizes brain dynamics as a “perpetual transient”. In dynamical systems theory, the term “transient” denotes the initial behavior of a system as it approaches an attractor. In the context of the brain, the dynamics can be said to be “perpetually” in this transient initial state, always approaching but never reaching a stable attractor (Friston, 1997). The reason why the stable attractor is never reached is because the state space of the attractor itself changes over time. These changes might for example due to neuromodulation: the same neural circuit can exhibit distinctly different modes of activity dependent on the degree of neuromodulation (Briggman & Kristan Jr, 2008; Engel et al., 2010).

Consider the following example to see how the two types of complexity relate. The dynamics of a neural circuit can be expressed by a set of differential equations:

$$\partial \boldsymbol{x}(t)/\partial t = f(x, \boldsymbol{C})$$

Given a set of state variables $\boldsymbol{x}$, and a set of (control) parameters $\boldsymbol{C}$, after the initial transient, the system comes to settle on its attractor. Now suppose a parameter $\boldsymbol{C}_1$ represents a control parameter internal to the system which itself varies over time as well (for instance neuromodulation through circadian rhythms). One can now attribute qualitative changes in dynamics as the consequence of a changing attractor manifold (type I) *or* one can integrate the changes of the control parameter $\boldsymbol{C}_1$ in the differential equations (type II), and can explain qualitatively different dynamics as a consequence of where on the attractor manifold the system is.

The idea of varying (control) parameters gives rise to the important idea of "attractors within attractors" (Friston, 2000a, p. 241): attractor dynamics over slow time-scales (for example that of circadian rhythms) constrain and modulate the attractor dynamics of faster time-scales (for example that of perception and action). We can therefore speak of a hierarchy of time-scales in the brain (Kiebel, Daunizeau, & Friston, 2008), where each time-scale provides the context for (or, more technically, modulates the attractor landscape of) the dynamics at a faster time-scale. Although the distinction between type I and type II complexity is ultimately a pragmatic decision (and it would therefore be a bad sign if one's philosophy of mind would change as a function of which perspective is adopted), taking a type II perspective does help to make some explanatory demarcations in a complex adaptive system like the brain.

It is important to note that the kind of dynamics that might modulate neurodynamics are not confined to the brain. For example, Ito et al. (2014) found that respiration modulates the neurodynamics in the neocortex of awake mice. This does not only occur in the olfactory bulb, but also in areas of the neocortex that are not involved in olfactory processing. Drawing upon Ito et al. (2014), Varga and Heck (2017) review a broader body of evidence that purports to show how respiratory dynamics not only modulates neurodynamics, but go on to show how respiratory dynamics might be functional for a wide range of cognitive processes. Varga and Heck conclude that:

> it is likely that respiration created an ever-present fundamental neuronal rhythm that may have shaped the temporal organization of neuronal activity during mammalian brain evolution. Some highly preserved spatiotemporal patterns of neuronal activity such as default mode networks (Greicius, Supekar, Menon, & Dougherty, 2009) might thus – in ways we do not yet fully understand – be the result of the incessant interaction between intrinsic network activity and respiration-locked neuronal rhythms. (p.83)

These findings show that the spatiotemporal organization of the brain's dynamics is contextualized and stabilized by the dynamics of the body. Hence, the respiratory pattern acts as a kind of "pacemaker" for the dynamics of the brain. To some theorists, this finding at least hints at the distinct possibility that, rather than that the brain enslaves the body, the body enslaves the brain (Dotov, 2014; Van Orden, Hollis, & Wallot, 2012). At least it shows the intertwinement of neural and extra-neural dynamics.

To conclude this section, the brain's dynamics are best characterized as a dynamical system with metastable attractor dynamics over multiple time-scales, each time-scale providing the context for dynamics at faster time-scales. Bodily dynamics, such as respiratory dynamics, can modulate neurodynamics as well. As such, the brain can be characterized as a highly labile organ in open contact with the dynamics of the body it is housed in and the world it inhabits.

Functionality: selection and anticipation

So far, only a *characterization* of (neuro-)dynamics was provided. The important question now is what its *functionality* might be. Put in other words: what are the selective pressures at work on the dynamics of the brain when put in the context of exchanges with the environment? Remember that on Freeman's account, the cortex can be seen as an attractor landscape. Taking a type II complexity perspective on this attractor landscape, there are dynamics over slower-time scales (presumably driven by sub-cortical structures, see for example Cisek & Kalaska, 2010; van Maanen et al., 2011) that act as control variables for the attractor landscape of the cortex. We can now put together a number of ideas presented in this introduction to understand the selective pressures at work on neurodynamics. First, responsiveness to affordances can be better or worse. Second, the attractor dynamics of the cortex structures the relevance of an agent's affordances, this is Freeman's model. Third, following Friston in taking a type II perspective, the attractor dynamics of the cortex is underpinned by dynamical control variables. Now, the extra assumption is that dynamical control variables are subject to selective consolidation. That is to say, the brain needs to be sensitive to when things are going right, and consolidate or 'reinforce' those dynamics, this is called value-based selection (Friston, Tononi, Reeke, Sporns, & Edelman, 1994). Combined, these assumptions lead to the idea that there are selective pressures on the intrinsic dynamics of the brain to steer the agent towards better interactions with the environment.

Besides this first-order selective pressure for dynamic control variables to steer the agent towards better interactions with the environment, there are second-order selection pressures at work as well. Like natural selection, neural selection thrives on diversity. Natural selection selects for the genes of an agent with the highest fitness, but this selection is contingent upon the right amount of variation. If no mutations in DNA-replication would take place, evolution is not possible, if too many mutations take place, stable evolution will not be possible as well. So, whereas first order selection selects for fitness, second order selection selects for *evolvability* (Partridge & Barton, 2000). Analogously, first order neural selection selects for those dynamic control variables that best fit with the situation. While second order neural selection, selects for *diversity* and *switchability.* This suggests, according to Friston that the source of diversity in the brain, metastable dynamics, it itself selected for: "If it is necessary to have metastability to facilitate neuronal selection then that metastability has, by definition, adaptive value. It will therefore be selected for at both an evolutionary and neuronal level" (Friston, 2000a, p. 248).

What is required for this story to work is a consolidation of 'successful' dynamic control parameters. This raises the question about what normative terms like'successful' and'right' mean in this context. One can think of 'getting things right' in an epistemic sense (i.e., having predicted what happens next) and 'getting things right' in a value-based sense (i.e., being in a flourishing state). In the engagements of an animal with its niche, these two questions are always intertwined: only by navigating its environment does the rabbit find food. From the perspective of evolution, there seems to be no reason to separate these two questions. Exactly how to understand normative terminology in the context of the dynamics of the brain, and how to understand the integration of value with epistemics, will be explored in **Chapter 2**.

The free-energy principle

The main aim of this section was to show the continuity between Freeman's neurodynamics (1987, 2000), and early neurodynamic work by Friston (1997, 2000b, 2000a). The common ground between these neuroscientists seems to be a shared focus on the importance of metastable dynamics and dynamics over slower time-scales that contextualize dynamics over faster time-scales. Already in the early work by Friston one can see an interest for the selective pressures at work on neurodynamics when put in the context of perception and action. These ideas get extended and formalized in more recent work on what is called the *free-energy principle* (Friston & Stephan, 2007; Friston, 2010). The central idea (which will be revisited a lot later in the subsequent chapters) is that since the brain occupies an invariant set of states, it can be considered to be a random dynamical attractor. But according to Friston:

> [A]ny system that possesses a random dynamical attractor must possess a dynamics that can be expressed as a minimisation of variational free energy, where variational free energy is the quantity that is minimised during approximate Bayesian inference [...]. In short, any system with a random attracting set must behave as if it is making Bayesian inferences about external states, using sensory impressions upon its Markov blanket. This formulation is offered as an endorsement of Walter J. Freeman's insights into the nature of neuronal dynamics in the self-organizing brain. (Friston, 2017, p. 119)

This is a puzzle. On the one hand, Friston is fully in keep with self-organization and neurodynamics and endorses Freeman style neurodynamics. On the other hand, he emphasizes how notions like inference and representation naturally emerge from self-organization. The continuity I tried to sketch between the free-energy principle via Freeman and Dreyfus with Merleau-Ponty and Heidegger, could then perhaps just as well be construed between the free-energy principle and philosophers like Kant and Helmholtz, the proponents of the mediational worldview. However, if Dreyfus and Taylor's (2015) analysis is correct, the Helmholtzian and the Merleau-Pontian approaches towards the mind are fundamentally incompatible. Indeed, in his philosophical interpretation of predictive processing, Hohwy (2014) ticks all the boxes that make up the mediational worldview: an agent can *only* have knowledge of what passes through the evidentiary boundary (a Markov blanket), knowledge is characterized as beliefs that are only justified by reference to other beliefs and sensory input, and evidentiary boundaries provide the dualist sorting between worldly states and mental states. Contrary to this, as mentioned earlier in the introduction, part of this dissertation presents the argument that the free-energy principle is best understood in terms of ecological and enactive philosophy of mind, and not in the rationalist philosophy of mind it is standardly associated with.

Introducing the philosophical neighbors

This dissertation takes it philosophical starting point in the philosophical legacy of Martin Heidegger (1996/1927), Maurice Merleau-Ponty (1945/1962, 1942/1966), Ludwig Wittgenstein (1953) and Hans Jonas (1966/2001). It is, however, not to be understood as an exegetical project, but rather as an attempt to use this philosophical background to, both critically and constructively, engage with modern cognitive science. As such, I join a broader movement of more or less radical strands of embodied cognitive science that have sparked over the last decade. Unified in what they are against, they disagree considerably in the positive outlook they have in mind. An exact comparison of all these new recent positions would be a considerable project on its own (see Käufer and Chemero (2015) for a start). In this section I will position myself in the current literature with respect to my philosophically nearest neighbors. I should note that there is a group of neighbors that is so close (with in some cases even neighboring offices) that it is impossible to position myself with respect to them. They have worked with me on an ecological-enactive account of skilled intentionality and how it can be put to use. Their work is discussed and relied upon extensively throughout this thesis (De Haan, Rietveld, Stokhof, & Denys, 2013; Kiverstein & Rietveld, 2015; Rietveld, De Haan, & Denys, 2013; Rietveld & Kiverstein, 2014; Rietveld, Denys, & Van Westen, in press; van Dijk & Rietveld, 2017).

Radical Embodied Cognitive Science

One of the most empirically minded approaches is *Radical Embodied Cognitive Science* (Chemero, 2009). Chemero argues that there are at least a class of cognitive phenomena that are best explained without recourse to mechanisms and computation, but are to be explained in terms of dynamical systems theory instead. In order to achieve this, Chemero argues against those who claim that mental representations are a prerequisite for cognition and adaptive behavior (or against anyone who puts *a priori* constraints on how to do cognitive science). Rather than rejecting mental representations across the board, he argues for explanatory pluralism (Dale, Dietrich, & Chemero, 2009). Given this pluralism, Chemero develops a cognitive science in which Gibsonian ecological psychology (Gibson, 1966, 1979) can be a guide to discovery, and explanations can be given by providing the equations that specify the dynamics between agent and environment.

What is lacking in Chemero's original (2009) formulation, and in ecological psychology more generally, is an account of agency or the fact that affordances are not just *possibilities* for action but can also be *invitations* to act (Rietveld, 2008b; Withagen, de Poel, Araújo, & Pepping, 2012). The same affordance can, dependent on context, skill and current concerns, be experienced in many different ways. The need to give an account of intentionality and agency in terms of dynamical and self-organizing systems is starting to be addressed (see Silberstein & Chemero, 2011; Withagen, Araújo, & de Poel, 2017), the current dissertation, and especially **Chapter 1**, which was originally published (Bruineberg & Rietveld, 2014 (Chapter 1)) in a special issue of *Frontiers in Human Neuroscience* on *Radical Embodied Cognitive Neuroscience*, can be seen as contributing to that project.

Autopoietic Enactivism

A second close neighbor is autopoietic enactivism, first introduced in *The Embodied Mind* (Varela et al., 1991) and further developed by Di Paolo (2005) and Thompson (2007). One of the main aims of autopoietic enactivism is to trace the origins of meaning, significance and subjectivity to the organization of the organism. Its two main pillars are autopoiesis theory (Maturana & Varela, 1980) and Jonasian philosophy of biology (Jonas, 1966/2001).

To start with the first, Maturana and Varela define an autopoietic system:

> [A]s a network of processes of production (transformation and destruction) of components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it [...] as a concrete unity in space in which they (the components) exist by specifying the topological domain of its realization as such a network. (Maturana & Varela, 1980, p. 78)

In other words: an autopoietic system is a system that through interactions with the environment brings about the very conditions that are necessary for its own existence. The paradigmatic example of such a system is a living cell: a bounded system that through energetic and chemical interactions with the environment is able to sustain a network of metabolic reactions that produces the lipid cell wall that demarcates the system. Life is understood as an active process that maintains its own organization.

In its original formulation, autopoiesis theory was conceived of as a formalization of the organization of a living systems and its dependency on the environment. As Maturana and Varela (1987) write: "[a]s long as a unity does not enter into a destructive interaction with its environment, we *as observers* will necessarily see between the structure of the environment and that of the unity a compatibility or congruence" (p.99, my italics). Hence, autopoiesis is an all-or-nothing phenomenon. The later Varela (Varela et al., 1991; Weber & Varela, 2002) as well as Di Paolo (2005) and Thompson (2007) argued that a more normative and organism-based (rather than observer-based) conception of the congruence between agent and environment was needed in order to account for "such biological phenomena as stress, illness, fatigue, and health, as well as plasticity and adaptation more generally" (Thompson, 2007, p. 148). Di Paolo (2005) proposes the notion of "adaptivity"[11] to denote the capacity of an organism to be sensitive and adapt to how well it's faring in its interactions with the environment. Di Paolo notes that: "This capacity may result from the action of dedicated mechanisms or it may be an emergent aspect of specific ways of realizing autopoiesis" (Di Paolo, 2005, p. 438).

The second pillar on which autopoietic enactivism is based is Hans Jonas' philosophy of biology (1966/2001). The starting point of this philosophy is the *mind-life continuity thesis* or the claim that "the organic even on its lowest forms prefigures mind, and [...] mind even on its highest reaches remains part of the organic" (Jonas, 1966/2001, p. 1). Life, understood as an active process that through material exchanges with the environment maintains its own organization, stands in a particular relation to matter: "[i]t is never the same materially and yet persists as its same self, *by* not remaining the same matter" (p.76, italics in original). Life *can not* be defined in terms of matter, exactly because its material make-up, but not its form, is accidental and fleeting. But the independence of life with respect to matter is balanced by the necessity to keep exchanging matter with its environment: "the sovereignty of form with respect to its matter is also its subjection to the need of it, by the impossibility of its resting with any simultaneous sum of stuff with which it happens to coincide in an instant of time" (p.83-84). But in order to exchange matter, the organism must be directed to the source of matter, the environment: "life is turned outward and toward the world in a peculiar relatedness of dependence and possibility" (p.84, my italics).

---

[11] More precisely, Di Paolo defines adaptivity as:
"A system's capacity, in some circumstances, to regulate its states and its relation to the environment with the result that, if the states are sufficiently close to the boundary of viability,

1. Tendencies are distinguished and acted upon depending on whether the states will approach or recede from the boundary and, as a consequence,
2. Tendencies of the first kind are moved closer to or transformed into tendencies of the second and so future states are prevented from reaching the boundary with an outward velocity." (2005, p. 438)

For Jonas, metabolism, purposiveness and significance arrive at the stage together at the origin of life, and they remain on stage, in all their diverse forms and developments, throughout the play. Note though that Jonas speaks of mind-life *continuity* and not *equality.* As Colombetti rightfully notes:

> mind shares the organizational properties of life, and richer forms of mind depend on richer forms of life. [...] But this means [...] not that consciousness, not even in some minimal nonreflective form, is present in all forms of life. Rather, the idea here is that the autonomous and adaptive organization of living systems sets up an *asymmetry* between them and the rest of the world, such that living systems realize a perspective or point of view from which the world acquires meaning for them, and not vice versa (Colombetti, 2014, pp.19-20).

The project of autopoietic enactivism is then to trace the evolution and development of the organization of life and mind in all its diversity, social and cultural embeddedness, complexity and pathology (Barandiaran & Moreno, 2006; Colombetti, 2014; De Jaegher & Di Paolo, 2007; De Haan, n.d.; Di Paolo, Buhrmann, & Barandiaran, 2017; Thompson, 2007).

Radical Enactivism

One slightly less empirically minded neighbor is radical enactivism (Hutto & Myin, 2013, 2017). Hutto and Myin want to make precise (and, indeed, radicalize) enactivist's claims that cognition does not involve the manipulation, extraction, processing, storage or pick-up of representations. They launch a full-scale attack on what they take to be the Achilles heel of the representational theory of mind: the notion of content to which any theory of representation is committed. They claim there to be representational content "wherever there are specified conditions of satisfaction"(Hutto & Myin, 2013, p. x).[12] The pincer movement with which Hutto and Myin approach the issue is to on one flank argue that, as a naturalistically inclined philosopher, one *can't have* conditions of satisfaction without embedding in the appropriate practices, and on the other flank argue that you *don't need* content in order to explain basic cognitive phenomena. After the battle is won, they enforce a peace treaty in which content is strictly forbidden for any basic mind unless in contact with the appropriate socio-cultural and linguistic practices.

There are a number of arguments and aspects of Hutto and Myin's proposal that are pertinent to the current thesis. They are allies in arguing against the prevalent intellectualism in cognitive science by setting the standards which any

---

[12] I should be quick to point out that they do not deny, or are concerned with, the phenomenal content or what-it-is-likeness of a state (Hutto & Myin, 2017, p. 11).

theory of content (which, I take it, is presupposed by any notion of desire, or belief) must meet. I will focus on one important dichotomy they introduce. The first pertains to the difference between *information-as-covariance* and *information-as-content*. Information-as-covariance is a philosophically innocent notion of information that pertains to the way things tend to correlate with one another: the number of rings in a trunk correlate with the age of a tree, my genes correlate with that of my siblings. This type of information is captured by the concept of *Shannon information*.

The notion of *information-as-covariance* should not be confused or conflated with the notion of *information-as-content* . While the former comes for free in the natural world, the latter is more philosophically expensive, having properties like truth, reference and implication. Hutto and Myin make clear that, and I think everyone would agree, mere covariance does not constitute content. The same holds for covariances between states of which one is inside and one is outside an organism, this very fact alone does not change the picture.

One the one hand here Hutto and Myin employ their *can't have* strategy, arguing against attempts to provide a naturalistic account of content, the most influential being Dretske's teleofunctionalist and Millikan's teleosemantic account. I will not reiterate their concerns here (see Chapter 4 of Hutto and Myin, 2013). On the other hand, and more interestingly for current purposes, they employ their *don't need* strategy. They want to retain an important aspect of Millikan's theory, namely the fact that through natural selection systems can become directed at specific aspects of their environment, but cash this out in non-contentful terms:

This is to accept that organisms often act successfully by making appropriate responses to objects or states of affairs in ways that are only mediated by their sensitive responding to natural signs, where this responding does not involve contentfully representing the objects or states of affairs in question (Hutto & Myin, 2013, p. 81).

In other words, in Hutto and Myin's eyes, the only naturalistic way forward is to accept non-contentful, selective directedness of an organism to aspects of its environment. Our routes converge here, since it is exactly such a directedness that I attempt to articulate in this dissertation. However, the positive account of directedness that radical enactivism offers (Hutto & Satne, 2015) remains focused on the selection of evolution of single "targeted response tendencies" and ignores the *organization* of response tendencies (action-readiness) that is basic to skilled intentionality (Kiverstein & Rietveld, 2015).

Philosophical methodology

The main aim of this dissertation is to articulate a *conceptual framework* that is able to make intelligible how the different aspects of skilled intentionality, as mentioned in the opening paragraphs, fit together. In this section I will make explicit what kind of methodology is employed in developing such a framework. There are two hard questions I have received, on multiple occasions, during my PhD (from philosophers and scientists, of the more mono-disciplinary variety): if not reading and interpreting classical texts, conducting experiments, analyzing data, developing mathematical models and forming testable hypotheses, then what *have* you been doing? And, closely related but not less vicious, if none of the above, then *what is it good for*? This is my best effort to provide an answer to these two hard questions.

What is a conceptual framework?

A "conceptual framework" is a tool to make explicit and intelligible the concepts and structures that constitute, underlie or in some other sense relate to a particular phenomenon. For example, Pacherie (2008) in a paper on the phenomenology of action aims "to propose a conceptual framework allowing for a *more precise characterization* of the *many facets* of the phenomenology of action, of their *relations to one another*, and of their *possible sources*" (p.180, my italics). Pacherie consequently introduces a "*key assumption guiding* this attempt" (p.180, my italics), in her case the assumption that the phenomenology and control of action are tightly interlinked, and goes on to develop a "dynamic model of intentions and action specification" (p.180) which can account for action control and the phenomenology of agency, hoping to show that this theory "provides a *framework for thinking* about action that is both *conceptually* and *empirically* motivated" (pp. 180-181, my italics).

Albeit brief, Pacherie nicely shows what the main desiderata of a conceptual framework are: it should be able to give a precise and rich characterization of a phenomenon, show how different aspects of the phenomenon fit together, what these aspects presuppose, and how they can be grounded. Furthermore, one can introduce a number of assumptions that help to guide to construct the framework. The proof of the pudding, for any conceptual framework I take it, is whether it is *internally consistent* and whether it *helps to think* about the phenomenon in question. The extra desideratum for frameworks at the interface between different disciplines and branches of science, such as both Pacherie's and the framework presented in this dissertation, is that they are able to *translate* between these branches and disciplines. This involves the shaping and molding of concepts such that they are able to be put to work in different disciplines. In the rest of this section, I want to focus on a number of specific aspects pertaining to conceptual frameworks: the role of assumptions in conceptual frameworks, how conceptual frameworks relate to theories, how they relate to worldviews, and how to do conceptual analysis in practice.

The role of assumptions

Everyone, even philosophers, needs assumptions. There is no shame in having them. Good assumptions reduce the space of possibilities of how the particular pieces of a conceptual puzzle can fit together: they act as constraints that guide theorizing. Assumptions come in many forms: sometimes explicit like in the case of Pacherie above, sometimes they are taken to be known to the audience, like when an author says "we take a Heideggerean approach", and sometimes the assumptions are inconspicuous to the author. I take it that one of the central jobs of philosophers is to make explicit what the assumptions are that are at work in a particular kind of conceptual framework.

Assumptions come at a price, and some assumptions are more expensive than others. Let me take one example of an assumption that figures prominently in the current dissertation. In **Chapter 1**, I write:"[w]e call this [phenomenon] the tendency towards an optimal grip on the situation, which we take to be a basic concern of every living animal" (p.43). That is to say, in order to understand skilled intentionality, one needs to *assume* the phenomenon of the *tendency towards an optimal grip* to be a basic organizational feature of living systems. How expensive this assumption is, depends on the specific work it is doing.

Dreyfus proposes a *structural isomorphism* (Dreyfus, 2007, p. 259) between brain dynamics and Merleau-Ponty's notion of the *tendency towards an optimal grip*. As said earlier in this introduction, it would be a costly assumption to hold that the phenomenological concept of optimal grip *explains why* brain dynamics is structured the way it is. This would seem like a brute metaphysical posit that introduces a vitalistic kind of teleology to otherwise mechanistic systems, then it is certainly extremely expensive. If, on the other hand, living systems are defined as systems that manage their interactions with the environment in a way that enforces their own conditions of viability, then optimal grip does not *ground* purposiveness, but *characterizes* it, and the assumption becomes much cheaper. I take it that in that case, the assumption is that an organizational feature of living systems is isomorphic with a particular structure of our experience.

Phenomenology and naturalism

There is one question that keeps creeping up when trying to integrate phenomenology and modern cognitive science: how to deal with the fact that phenomenology rejects, and cognitive science seems to presuppose, naturalism? The (im)possibility of the naturalization of phenomenology is, of course, a heavily debated topic of its own (Gallagher & Zahavi, 2007; Kiverstein, 2006; Petitot, 1999). I will not delve into these debates here, but only briefly point out my commitments. A first thing to note is that *if* naturalism means the ultimate *reducibility* of meaning and intentionality into merely causal interactions, then it should be rejected. As Thompson writes in *Mind in Life* (Thompson, 2007, p.

81): "[N]aturalism cannot explain matter, life and mind, as long as explanation means purging nature of subjectivity and then trying to reconstitute subjectivity out of nature thus purged."

However, different models of naturalizing phenomenology are available. For example, Silberstein and Chemero (2012) simply assume the inseparability of meaningful cognition and consciousness: "experience is cognition and cognition is experiential. Our cognitive, conscious, and behavioral capacities co-explain and co-determine each other dynamically" (p.40-41). Now, this is an assumption, but it has its own distinct advantages. For one, there is no need to posit qualia and think about how they relate to causal relations in the world. Rather, the problem of consciousness becomes equivalent to the problem of intentionality. Ecological, dynamic and enactive approaches to cognition *do* have the tools to study that problem. As Silberstein and Chemero (2012) write:

> Extended phenomenological-cognitive science has all the advantages of the identity theory, according to which conscious experiences are identical to brain states, without the problem of explaining how conscious experiences such as seeing red could be Nothing But neurochemical events; the phenomenological world of experience is neither in the head nor in the external world - it is fundamentally relational. (Silberstein & Chemero, 2012, p. 41)

This is an assumption in exactly the sense mentioned in the previous paragraph: it constrains and guides theorizing about consciousness, without itself being justified. Whether this assumption comes at an acceptable price depends partially on the cost of alternative assumptions. Alva Noe, defending, I take it, a similar form of phenomenological realism, points to the underprized assumptions of the internalist competition:

> How you get from a "mental model" [in the head] to the experience of the world is one of the biggest IOUs scientists have ever tried to pass off on unsuspecting graduate students. (Noe, 2012, p. 30)

However, one might at this point wonder whether this still is an interesting form of naturalism. In fact, in a more recent paper, Silberstein (2015) indicate that their account of *Extended Phenomenological-Cognitive Systems* is best paired with neutral monism. These more metaphysical aspects reach beyond the topic of this dissertation. What matters here, is that the position allows for the use of tools from the natural sciences, such as dynamical systems theory, to study behavior, cognition and consciousness.

Conceptual analysis: inference and resonance

The conceptual framework proposed in this dissertation, the skilled intentionality framework, spans many different fields. One of the main dangers of such a multidisciplinary approach is that concepts used in one field do not translate neatly to another field. Dangerous cases include situations where the same term is used in different fields, such as the use of *inference* in computational neuroscience and philosophy of mind, and where the original term comes to mean something quite different, such as in the use of the term *affordance* in the domain of neuroscience. What is required then is careful analysis of the concepts in question, the exact work they are doing within a particular field in order to acquire an understanding of the possibilities and limitations of the use of a concept in another. I will take the aforementioned example of *inference* as an illustration of this conceptual work. I will discuss the notion of inference in the context of Jakob Hohwy's argument a commitment to predictive processing automatically leads to inferential seclusion and Cartesian skepticism. In other words, a commitment to exactly the picture of the mind that Dreyfus and Taylor (2015), and this dissertation, argue against.

I take there to be (at least) two stable positions in the philosophy of perception that are mutually exclusive and that each are not trivially wrong. One is direct realism, roughly the view that, generally speaking, we come in direct unmediated contact with the relevant aspects of our environment. The other is indirect realism or representationalism, roughly the view that, since perceptual input is impoverished, we need to internally reconstruct our environments based on sensory input. The content of our experience is not the world out there, but rather our best estimate, or internal hypothesis about the worldly causes of our sensory information. This is exactly the opposition between the *mediational picture* and the *contact theory* laid out by Dreyfus and Taylor (2015).

These two opposing positions show up again in the psychology of perception of the second half of the 20th century. Two pioneers in the field each developed a research program based on these two opposing directions, each modeling their view of perception after a specific analogy. Richard Gregory (1980, 1997) developed a view of *perception as hypothesis-testing*, taking the perceptual system to function analogous to the scientific method: sensory input is taken to be data, which can then serve as evidence for a particular hypothesis. The brain then comes to settle, through a process of testing and adapting hypothesis to settle on the most likely hypothesis given the data and prior knowledge. On this picture, the perceptual system is doing *inference to the best explanation*, where what is explained is the sensory input, and the best explanation of that input is the worldly state of affairs that most likely generated that input.

Gregory (1997) contrasted his ideas explicitly with that of another pioneer in the psychology of perception: James J. Gibson. In his book *The Senses Considered as Perceptual Systems* (1966), Gibson introduces the term *resonance* to describe the relation between perceptual systems and their environment:

> Instead of supposing that the brain constructs or computes the objective information from a kaleidoscopic inflow of sensations, we may suppose that the orienting of the organs of perception is governed by the brain so thatthe whole system of input and output resonates to the external information (Gibson, 1966, p. 5).

The function of the brain, according to this view, is to *resonate to* ecological information. Now, as many authors have noted already (Teques, Araújo, Seifert, del Campo, & Davids, 2017; Raja, 2017), the notion of resonance is underdeveloped in own Gibson's work. However, I take it that since Gibson's work a fully-fledged dynamicist framework for cognitive science has emerged, revolving around notions like *resonance*, *synchronization*, *entrainment* and *coordination* (Stepp, Chemero, & Turvey, 2011).

The latter resonance-based view to perception is also known as *direct* perception. The former, inference-based view of perception, mediated by sensory input is also known as *indirect* perception. The point here is that *resonance* and *inference* are irreducibly different concepts stemming from completely different philosophical traditions, figuring in completely different conceptual frameworks. Still, on a very abstract level there are some similarities. Most notably: both concepts involve a relation between (at least) two processes. In the case of *inference* one process *infers* the state of another process based on the input it is getting. In the case of *resonance* two systems are *resonating* through being coupled to one another. Another similarity is that both concepts work with some sort of measure of proximity over time: the inferring system is over time *approaching* the best explanation. The two coupled systems over time *synchronize* or *attune* to each other. The main difference, I take it, is that *inference* is committed to a notion of content. The inferring process is holding, in whatever kind of sense, a hypothesis about the inferred process. This is not the case for *resonance.*

Now that we have done the conceptual groundwork and have the relevant concepts and contrast classes on the table, we can take a look at the structure of Hohwy's argument for indirect perception and inferential seclusion based on predictive processing. Let's start with his account of internalism:

> It is tempting to say that any account of perception and cognition that operates with internal models must in some sense be internalist. But the natural next question is what makes internal models internal? I think a natural default has been to answer that internal models are internal because they are housed in the brain. But this cannot be right. [...] A better answer is provided by the notion of Markov blankets and self-evidencing through approximation to Bayesian inference. Here there is a principled distinction between the internal, known causes as they are inferred by the model and the external, hidden causes on the other side of the Markov blanket. This seems

> a clear way to define internalism as a view of the mind according to which perceptual and cognitive processing all happen within the internal model, or, equivalently, within the Markov blanket. This is then what non-internalist views must deny (Hohwy, 2017, pp. 6-7).

So, according to Hohwy, all the heavy philosophical work to argue for internalism is done by the concept of a Markov blanket. They "screen off" the world from the mind: the mind can only make reference to whatever information passes through the Markov blanket. Now, for the current argument it does not actually matter what Markov blankets are,[13] it just suffices that the claim is that anyone who accepts that notion will, according to Hohwy, be an internalist.

As Hohwy shows in the same paper, one example of a system with a Markov blanket is Huygens pair of pendulum clocks (Huygens, 1673). He insists that, since this system has a Markov blanket, these clocks are inferentially secluded, and in fact holds that one clock is inferring the state of the other clock.[14] The problem is that this very same example is what direct-perception theorists have in mind when talking about direct perception as *resonance*. It seems we find ourselves in an impasse.

Are we to follow Hohwy in believing that all resonance actually is inference? Should we dismiss Markov blankets as a demarcation criterion? And if so, by which standards? What can arbitrate in such a kind of dispute? Now, I take it there is no way to empirically settle the issue: we *can't test* whether coupled clocks do inference or not, we need to find a solution in the conceptual domain.

Relating the notions of *inference* and *resonance* with their embeddedness in different conceptual frameworks in the psychology of perception and, ultimately, in the philosophy of mind might be helpful. For, as Hohwy notes himself: "this definition of internalism makes Clark an internalist" (p.7), as, I should add, it would make Gibson, myself, and any *resonance*-based theorist internalists. Furthermore, it would make utter nonsense of the claim that we can be in direct contact with the world since this would require *unmediated causal contact* (of the sort bullets deliver). We can now see more clearly what is at stake: accepting Hohwy's demarcation criterion collapses the distinction between Gibson and Gregory's theories of perception, and ultimately collapses the distinction between a contact-based and a mediational mind-world relationship. The cost of making non-sense of externalism is trivializing internalism.

Using Dreyfus and Taylor (2015) we can give a crystal-clear diagnosis of what actually is going on. In their analysis, boundaries are only one of four strands that together support the internalist world-view. If you implicitly accept the rest

---

[13] Markov blankets pertain to particular conditional independencies in a causal model (for those unfamiliar with Markov blankets, see **Figure 2.2**) for an explanation)

[14] At a workshop in Frankfurt in 2016, I raised this example of the coupled clocks as a clear *reductio ad absurdum* to his Markov blanket demarcation criterion for internalism. He seemed not impressed and happily included the example in the subsequent paper.

of the topology in which knowledge consists of internal states and justification can't go beyond these internal states, then of course one will find boundaries to be *epistemic* boundaries. Hohwy thinks it is Markov blankets that in themselves demarcate internalism from externalism, but he actually is presupposing the rest of the internalist topology to begin with. Externalists don't have to deny Markov blankets, they just deny that knowledge is in the head in the first place.

Perhaps the most troubling aspect about the whole situation is that Hohwy relies on a dominant use of the term inference in neuroscience (see for example Dayan & Hinton, 1996). It has been my own personal experience as well in talking to neuroscientists that the mere proposal that the brain might not have to be inferentially secluded from the world is met with absolute dismay. Again, here one might think that neuroscience is at the brink of discovering the truth of internalism, *or* that the neuroscientists and philosophers employ some very different understandings of inference. While Hohwy thinks the former, I choose the latter.

I think there is no ground, and no need, to say that neuroscientists are *wrong* to employ the concept of inference in neuroscience. However, as I will discuss in **Section "Zooming out"** this does not mean that therefore the concept of inference as used in neuroscience is directly applicable to debates in philosophy of mind. The claim that no "philosophical processing" (Godfrey-Smith, 2001) of neuroscientific concepts is necessary, is highly contentious and by no means the standard option. I think it is highly problematic to apply concepts from neuroscience to philosophy of mind without being sensitive to the different conceptual frameworks in which they figure and without carefully analyzing the scope and limitations of the concepts involved. In this dissertation, I try to do this as good as I can, although, I will admit, it is always easier to spot the mistakes in someone else's argument than in one's own.

Zooming in: from framework to theories

Frameworks should not be confused with theories, and different normative criteria are applicable to each of them: a good theory makes falsifiable empirical predictions. A good framework is internally consistent, helps one to think, and, at least in the context of science, helps one to articulate theories.

Accusing the free-energy principle - channeling our inner Poppers - of being unfalsifiable, is therefore ill-founded. But the free-energy principle *should* be able to inform theories, and one of the most promising aspects of it, is that it is able to do so in multiple scientific domains. One can roughly discern three domains in which the free-energy principle can be applied: the neural, the cognitive and the ecological. In the neural domain, the differential equations that follow from free-energy minimization (see **Chapter 5, Table 2**) can, in principle, be directly mapped onto neuronal dynamics (see Friston, Rosch, Parr, Price, & Bowman, 2017).[15] In the cognitive domain, the free-energy principle proposes a distinct integration of perception, action and attention. This allows one to operationalize, disambiguate and give a precise definition of cognitive concepts, like for example *attention* and *salience* (Parr & Friston, 2017). In the ecological domain, the free-energy principle works with a strong kind of mutuality between agent and environment (see quote by Friston on p. 53 and **Chapter 2** of this dissertation).

When focusing on a single one of these domains, the yields of taking a free-energy principle perspective can seem a bit meagre. For example, one can show that reinforcement learning is a special case of free-energy minimization (Friston, Daunizeau, & Kiebel, 2009; Friston, Samothrakis, & Montague, 2012). The central point of those papers is to show how *value* can be understood in terms of priors and expectations and does not need to be posited as a primitive of the theory. Importantly, these papers do *not* correct reinforcement learning in any way, but in fact go through great lengths to show that some paradigmatic examples in reinforcement learning can be modeled using the mathematical tools and concepts of the free-energy principle. The free-energy principle seems therefore vulnerable to a charge that is akin to non-falsifiability, namely that it, as a framework, *does not constrain theorizing*, it does not make a difference for a domain like reinforcement learning. This might seem true when focusing on only one domain. But given that the free-energy principle is able to connect to the neural, the cognitive and ecological domains and different theories within these domains, its main promise is that will be able to translate insights from one domain and use these insights as constraints on theorizing in another domain. Whether the free-energy principle succeeds as a unifying framework will be dependent on its ability translate between domains and to constrain and guide theorizing.

Zooming out: from framework to worldview

As we know since Kuhn (1962), conceptual frameworks have ontological implications. That is to say, they imply a worldview. In the context of the study of the mind, these worldviews can have strong epistemic, ethical and existential consequences. There is much more at stake than simply an argument over the most coherent philosophical interpretation of scientific facts. To return to the dichotomy from earlier this introduction, Dreyfus and Taylor state that:

> [t]he battle between the two construals, mediational and contact, is far from being a bloodless debate over scientific method. It is deeply involved in the contrary ethical and metaphysical passions of the modern age. (Dreyfus & Taylor, 2015, p. 26)

On the standard mediational picture: "The apricot-pink of the setting sun is not a property of the evening sky; it is a property of the internal model of the evening sky, a model created by your brain" (Metzinger, 2009, p. 20). But what

---

[15] However, this is not the topic of this dissertation.

holds for colors, holds just as well for all perceptual objects: "the seen rose, the touched icecube, the heard melody, the colleague, the group of refugees, they are all internal to and contained in the brain" (Zahavi, 2017). It seems highly dubious to me that one can hold that *all* we experience is a figment of the brain without this having existential and ethical consequences, for example, for our understanding of ourselves and our relationships with others. Note that the worry is not so much that the external world, including our loved ones, might not be there, but that we are in an epistemic position in which they are, forever, off-limits.

Should we pliably accept the ethical, epistemological and existential consequences of our best scientific theories of the mind or should we reject those theories of the mind exactly because of their ethical, epistemological and existential consequences? It seems that either option is highly unattractive. Perhaps a way out is offered by Shaun Gallagher (2017). In discussing the ecological-enactive interpretation of the free-energy principle offered in **Chapter 2**, Gallagher suggests that, borrowing a distinction from Godfrey-Smith (2001), some of this work might be better seen as offering a'philosophy of nature' rather than a'scientific research program'. Godfrey-Smith defines a philosophy of nature as:

> [A]n attempt to describe the world in a way that is closely informed by scientific theories, but that is free to reject the vocabulary and perhaps some of the classifications and interpretations of the world associated with the relevant sciences (2001, p. 286).

How much "philosophical processing" scientific theories need is dependent on one's view about how science works, and the status of scientific terms. If one for example holds that, in order to function properly, ordinary scientific work in a particular domain requires the continuous deployment of a rich set of metaphors, then part of the job of the "philosopher of nature" will be to weed through the scientific descriptions and sort the descriptions that are to be taken literally from those that are to be taken metaphorically (Godfrey-Smith, 2001).

Now, one should not be too modest here: philosophical processing of science *can* and *should* feed back into the science itself, since a better conceptual framework should lead to better theorizing, and hence to better hypotheses and better empirical research. But there are reasons to be pessimistic as well. Godfrey-Smith discusses the use of terms like *transcription, translation,* and *proof-reading* in genetics, a set of metaphors that he himself opposes.

> Maybe the symbolic perspective [. . .] just happens to be a uniquely useful framework for guiding empirical work in this area. If that is true, it is true not just because of what genes are like, but also because of what our minds are like. A conceptual framework like this might, for a mixture of reasons, turn out to be ideal for human rumination and

> communication about genetics. That could be true even if genes have many properties that are not captured by the symbolic perspective (Godfrey-Smith, 2001, p. 287).

At the same time, the semantic perspective on genes is not only pragmatic, "when used in nontechnical discussions in newspapers and the like, [it gives] support to genetic determinist views widespread in the general population" (Godfrey-Smith, 2001, p. 288).

One aspect of the work in this dissertation is therefore to "philosophically process" the vocabulary of the predominant framework in computational neuroscience, which revolves around terms like *representation*, *belief* and *inference.* Clark (2016b) seems to do something similar. When discussing the radical edges of predictive processing, Clark writes: "Could we perhaps have told our story in entirely non-representational terms, without invoking the concept of a hierarchical probabilistic generative model at all? [...] [A]s things stand, I simply do not see how this is to be achieved" (p. 293). At the same time, Clark wishes to reject the two main philosophical implications of representationalism: internalism and intellectualism. One way to understand this is take Clark to accept the scientific indispensability of representational language, but to simultaneously hold that there is nothing here that can not be suitably "philosophically processed". This is a delicate balancing act: too much "philosophical processing" and the relevance of science for philosophy is lost (everything one could have learned from science has been processed away), not enough "philosophical processing" and does not get rid of problematic implications. This balancing act shows up in Clark's use of "*not in-direct perception*" as a way of being able to accept the inferential language, but to also still be able to maintain some form of externalism. As Anderson (2017) notes: "[t]he double negative speaks volumes about Clark's conceptual struggle, here" (p.10).

In this dissertation, I take a different route. While, Clark (2016b) wants to save externalism by appealing to the way predictive processing is action-oriented and non-reconstructive , I use Dreyfus' "philosophical processing" of Freeman's neurodynamics as a starting point and develop both the continuity between Dreyfus' account of intentionality and *skilled intentionality,* and the continuity between Freeman's and Friston's accounts of neurodynamics. The result is an alternative conceptual framework for the anticipating brain that is not making use of terms like *representation*, *belief* and *inference.* I think this framework is both able to capture properties of the anticipating brain that the semantic perspective is less well able to capture, like metastability, and is able to provide what Dreyfus and Taylor call a "contact theory" of intentionality.

Summary of chapters

The chapters of this dissertation are written as independently readable papers and can be read in any order. The philosophically inclined reader might stick to the order as I present them here. The mathematically inclined reader might wish to start with **Chapter 5**, which is the most formal presentation of the free-energy principle and the work it can do.

- **Chapter 1** develops a conceptual framework for *Radical Embodied Cognitive Neuroscience.* The first part of the chapter introduces skilled intentionality and the structure of the landscape of affordances. The second part of the chapter relates skilled intentionality to theories of self-organization and neurodynamics, such as the free-energy principle. The third part of the chapter exemplifies the integrative approach by presenting research on human movement science and the impact of Deep Brain Stimulation on affordance responsiveness.

- **Chapter 2** develops the link between ecological-enactive cognitive science and the free-energy principle. The first part of the paper presents the free-energy principle and its commitment to the mind-life continuity thesis. In the second part of the chapter, the "crooked-scientist" metaphor is developed as a way to make sense of the integration of value and epistemics under the free-energy principle. In the third part of the chapter, the link between internalism, inference and Markov blankets is challenged by discussing coupled clocks as an intuition pump

- **Chapter 3** compares a rationalist Helmholtzian, a cybernetic and an ecological-enactive interpretation of predictive processing. In the second part of the chapter, the discussion focusses on how each of these three interpretations conceives of the sense of agency and intentionality in different ways.

- **Chapter 4** focusses on the structure of the socio-material environment. In particular, it presents a novel account of ecological information designed to work for cases of'higher' cognition. Introducing the notion of *general ecological information*, an account is given of these regularities in terms of constraints, information and the form of life or ecological niche.

- **Chapter 5** uses computational simulations of a free-energy minimizing agent to show that free-energy is a relational quantity, pertaining to the'fit' between an embodied agent and its econiche. A formal similarity between the way an agent remembers its environment and the way the environment'remembers' the behavior of an agent is exploited, and it is shown how niche construction is critically dependent on the learning rate of the agent, and the'inertia' or malleability of the environment.

Author contributions

The chapters in this dissertation have either been published in article form, or are currently in preparation. Below I will provide, if existing, the reference to the published version, if co-authored, note the relative author contributions.

- Chapter 1 has previously been published as:

  Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in human neuroscience,* 8 , 599. doi: 10.3389/fnhum.2014.00599

  Author contribution: JB and ER co-wrote the paper building upon ER's VIDI proposal on Skilled Intentionality.

- Chapter 2 has previously been published as:

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

  Author contribution: JB, JK and ER conceived of the paper together with JB constructing the core arguments of the paper. JB and JK co-wrote some parts of the paper with revisions from ER.

- Chapter 3 has previously been published as:

  Bruineberg, J. (2017). Active inference and the primacy of the'I Can'. In T. Metzinger & W. Wiese (Eds.). *Philosophy and Predictive Processing*: 5. Frankfurt am Main: MIND Group.

- Chapter 4 is currently in press at *Synthese* and has Bruineberg, Chemero and Rietveld as its authors.

  Author contribution: JB, AC and ER conceived of the paper together. JB wrote the paper with revisions from AC and ER.

- Chapter 5 is in preparation and has Bruineberg, Rietveld, Parr, van Maanen and Friston as its authors.

  Author contribution: Conceived and designed the simulations: JB, KF, TP, ER, LvM. Ran the simulations and analyzed the results: JB, TP. Wrote the paper: JB, KF with revisions from TP, ER, LvM.

Chapter 1

Self-organization, free-energy minimization and optimal grip on a field of affordances[1]

Abstract

In this paper, we set out to develop a theoretical and conceptual framework for the new field of *Radical Embodied Cognitive Neuroscience*. This framework should be able to integrate insights from several relevant disciplines: theory on embodied cognition, ecological psychology, phenomenology, dynamical systems theory, and neurodynamics. We suggest that the main task of Radical Embodied Cognitive Neuroscience is to investigate the phenomenon of skilled intentionality from the perspective of the self-organization of the brain-body-environment system, while doing justice to the phenomenology of skilled action. In previous work, we have characterized skilled intentionality as the organism's tendency toward an optimal grip on multiple relevant affordances simultaneously. Affordances are possibilities for action provided by the environment. In the first part of this paper, we introduce the notion of skilled intentionality and the phenomenon of responsiveness to a field of relevant affordances. Second, we use Friston's work on neurodynamics, but embed a very minimal version of his free-energy principle in the ecological niche of the animal. Thus amended, this principle is helpful for understanding the embeddedness of neurodynamics within the dynamics of the system "brain-body-landscape of affordances." Next, we show how we can use this adjusted principle to understand the neurodynamics of selective openness to the environment: interacting action-readiness patterns at multiple timescales contribute to the organism's selective openness to relevant affordances. In the final part of the paper, we emphasize the important role of metastable dynamics in both the brain

---

[1] This paper was previously published as Bruineberg and Rietveld (2014).

and the brain-body-environment system for adequate affordance-responsiveness. We exemplify our integrative approach by presenting research on the impact of *Deep Brain Stimulation* on affordance responsiveness of OCD patients.

1.1 Introduction

This *Frontiers* special issue on *Radical Embodied Cognitive Neuroscience* invites researchers to re-imagine cognitive neuroscience in terms of (radical) embodied cognitive science. Radical embodiment is the view that cognition ought to be understood primarily in terms of the embodied agent–environment dynamics. Neural dynamics can only be studied while taking into account the larger brain-body-environment dynamics (Chemero, 2009). Besides highlighting the dynamical aspects of cognition, embodied cognitive science has also highlighted the importance of phenomenology and ecological psychology for studying cognition. In this paper, we develop a theoretical and conceptual framework that aims to integrate some of the various fields of study that come together in a *Radical Embodied Cognitive Neuroscience*: neurodynamics, ecological psychology, phenomenology, self-organization and dynamical systems theory.

The starting point of this paper is the question how skilled agents interact with their environment and can tend toward improvement of their situation. In particular, we are interested in how, in a particular context, skilled agents are selectively responsive to only some of the many available "affordance" or possibilities for action offered by their environment (Gibson, 1979; Chemero, 2003). In order to understand this, phenomenology suggests that we need to complement Gibson's original theory of affordances with an understanding of the attracting or soliciting character of affordances in relation to an agent in a particular situation (Rietveld, 2008a; Withagen et al., 2012). We think that the main task of *Radical Embodied Cognitive Neuroscience* is to explain how the changing world and the dynamics of the agent's state mesh together in a way that makes adequate action possible, while simultaneously doing justice to the phenomenology of skilled action. In this paper we theoretically and conceptually develop a framework for investigating this. Although the phenomenon of skilled activity is relevant for both humans and non-human animals (Ingold, 2000), we will focus on human beings in this paper. Also, we will limit ourselves to agents who have already acquired their skills. So we will not focus on developing, learning, fine-tuning, and modifying skills nor on the evolutionary history of skilled behaviors, although these topics raise important open issues as well.

In the first part of this paper, we focus on the phenomenon of selective affordance-responsiveness because that is an ecologically valid way to characterize the dynamics of the system "skilled agent–environment". In the second part of the paper, we show how theoretical neuroscience can help to understand selective affordance-responsiveness. First, we introduce the framework of self-organization in order to bring the necessary conceptual tools to the table. Second, we focus on how neurodynamics is embedded in the dynamics of the broader brain-body-environment system. We present the free-energy principle (FEP) as a promising framework to understand this embeddedness but, inspired by Anderson and Chemero (2013), interpret it in a more minimal way than has previously been done. Furthermore, we show how we can use this adjusted framework to understand the neurodynamics of selective openness to affordances. Next, we argue for a situated understanding of the FEP in which the self-organizing brain is understood as coordinating action-readiness patterns to deal with relevant affordances. In the final part of the paper, we illustrate the plausibility of our conceptual framework by showing how it is able to integrate findings on metastable dynamics in the brain-body-environment system, and how it is able to shine new light on the effects of Deep Brain Stimulation (DBS) on treatment resistant obsessive-compulsive disorder (OCD).

1.2 Skilled intentionality and optimal grip on a field of affordances

Affordance-responsiveness is a central feature of everyday skillful activity of both humans and non-human animals (Rietveld, 2012a). Affordances are possibilities for action provided to an animal by the substances, surfaces, objects, and other living creatures that surround it (Gibson, 1979; Reed, 1996; Heft, 2001; Chemero, 2003, 2009; Silva, Garganta, Araújo, Davids, & Aguiar, 2013). Affordances can be defined as relations between aspects of the material environment and abilities available in a form of life (Rietveld & Kiverstein, 2014, cf. Chemero, 2003).

Up till now in the field of Embodied Embedded Cognition affordances have typically been understood as motor possibilities the environment offers to a creature, such as reaching, grasping, sitting, walking etc. Developing a Wittgensteinian account of affordances, we (Rietveld & Kiverstein, 2014) have argued that for creatures that inhabit a resourceful social and cultural environment as we do, the possibilities for action the environment offers are far richer: the affordances on offer in the landscape of affordances available in our form of life are related to the whole spectrum of abilities available in our human socio-cultural practices (cf. Heft, 2001). Both unreflective action in everyday life and episodes of what are traditionally called “higher” cognition are forms of *skilled* interaction with the environment and can be understood in terms of responsiveness to affordances (Rietveld, 2008c, 2013).

Based on a careful reading of Gibson, we have recently shown (Rietveld & Kiverstein, 2014), that contrary to what many think, it is not affordances but the ecological *niche* for a kind of animal with a particular way of life that forms the cornerstone of Gibson’s ideas. Our notion of the landscape of available affordances was introduced to do justice to this *primacy of the niche*, which is present

Affordance: A possibility for action provided by the environment to an animal.

Solicitation: An affordance that stands out as relevant for a particular animal in a specific situation.

Skilled intentionality: The kind of intentionality an individual exhibits when acting skillfully in a familiar situation (see main text for elaboration). We characterize skilled intentionality as the tendency toward an optimal grip on a field of affordances.

Tendency towards an optimal grip: The tendency of a skilled individual to be moved to improve its grip on the situation by responding to solicitations.

Landscape of affordances: The affordances available in an ecological niche. In our human form of life, these are related to the whole spectrum of abilities available in our socio-cultural practices.

Field of affordances: The affordances that stand out as relevant for a particular individual in a particular situation; i.e., the multiplicity of affordances that solicit the individual.

Table 1.1: Terminology of skilled intentionality

independently of perception by a particular individual (See **Table 1.1**). The astonishing richness of the landscape of available affordance in our niche hinges on the fact that both relata of affordances, both the sociomaterial environment and the reservoir of abilities in our socio-cultural practices, manifest an enormous variety.

This enormous richness raises the question how an organism can be responsive to only the *relevant* affordances in a given situation. Phenomenologically, some of the affordances around us do not leave us cold but move us. In earlier work (Rietveld, 2008a) we have suggested that an affordance can “invite” or “solicit” behavior dependent on the current concerns of the organism and the situation it is in (Withagen et al., 2012). The metaphor of a *field* is useful here: some affordances stand out more than others. Some are experienced as soliciting immediately, others are experienced as soliciting on the horizon and still others are completely ignored (only the latter do in fact leave us cold). We can distinguish between an affordance, i.e., a possibility for action available in our form of life at a certain location, and a solicitation. A solicitation is an affordance that stands out as relevant in a specific situation lived by an animal. “Action-readiness” (Frijda, 1986, 2007) is a useful notion here, because it is a phenomenon in between overt action and ability. A solicitation is the (pre-reflective) experiential equivalent of a bodily action-readiness: the readiness of the affordance-related ability (Rietveld, 2008a).

Much of our everyday interactions with the environment, such as riding a bike through a city, moving toward an appropriate distance from other people in an elevator, or ordering a cup of coffee in a bar, can be described as skillful activities. In previous work, we have introduced the notion of *skilled intentionality* as the tendency toward an optimal grip on a situation by being selectively responsive to available affordances (Rietveld, 2008c, 2012a, 2013). The tendency toward an optimal grip[2] is a primarily phenomenological notion that signifies the way a skilled individual acts in a familiar environment in order to improve its grip on the situation. What is central to this notion, is that the individual experiences the situation in terms of a deviation of an optimum. As Merleau-Ponty puts it:

> For each object, as for each picture in an art gallery, there is an optimum distance from which it requires to be seen, a direction viewed from which it vouchsafes most of itself: at a shorter or greater distance we have merely a perception blurred through excess or deficiency. We therefore tend toward the maximum of visibility, and seek a better focus as with a microscope (Merleau-Ponty, 1945/1962, p. 352).

Importantly, during those episodes of skilled activity, the skilled individual does not have an explicit goal in mind, but rather is solicited by the environment in such a way as to improve her grip on the situation. Phenomenologically, this deviation of an optimum can be described as an experienced tension to be reduced. In the case of a skilled individual, which is what we focus on in this paper, tending toward grip is the equivalent of having an action-readiness for dealing adequately with an affordance; one is responsive to, or poised to act adequately on an affordance.

We suggest that the tendency toward an optimal grip on the situation is a basic concern of living organisms and is a central feature of our everyday skillful dealings with our environment. It shapes the person's selective openness to the landscape of available affordances so that certain affordances "stand out" as relevant and the individual can unreflectively improve his or her situation by simply being responsive to this structured field of relevant affordances (Rietveld, 2008c, 2012a, 2012b). For instance, when entering a crowded elevator, we stand at an appropriate distance from the other people.

---

[2] The word grip has several connotations in the English language. It can refer to a physical grip (such as when grasping a cup), but also a more intellectual grip (such as when having a grip on a problem), as well as a grip in the sense of being able to deal with something (such as when losing your grip on a situation). As we state in the text, skilled action pertains to simple motor behaviors, but also to more complex and context-sensitive actions. Optimal grip is, because of the multiple connotations of the word grip, supposed to characterize all these aspects of the phenomenology of skilled action.

It is this phenomenon of the tendency toward an optimal grip and especially how theories from the fields of self-organization and theoretical neuroscience can contribute to an understanding of context-sensitive selective openness to relevant affordances that is the central topic of this paper.

The specific structure of the field of affordances of a particular individual is dependent on the current concerns and abilities of that organism and the current situation. The structure of the field of affordances changes when either the landscape of affordances changes (i.e., when the sociomaterial environment changes or when the abilities available in a form of life change), or when the concerns of the individual change. If a rabbit eats the only carrot available in a certain place, it changes the layout of the (locally present) landscape of affordances. However, as the landscape of affordances changes and the individual's interest in eating diminishes, new possibilities for action show up. Once the carrot has been eaten, the rabbit hole might solicit sleeping, or a place a bit further away might solicit exploring (cf. Dreyfus, 2007).

Changes in the field of affordances can also originate in the environment. For the eating rabbit, a sound in the bushes might change the field in such a way that the carrot does not solicit eating anymore, but now the rabbit hole solicits hiding. An important part of skilled intentionality is therefore not only being skillfully responsive to one affordance, but also being open to changes in the context and adequately engaging with these affordances (see also **Section 1.7** on metastability). The tendency toward an optimal grip on a field of affordances is the result of a dynamic interplay between the landscape of affordances and the current state of the organism. On the side of the organism, states of action-readiness interact in order to bring about selective openness to a landscape of affordances (see **Figure 1.1**). We will return to the processes of self-organization and neurodynamics contributing to selective openness in the subsequent sections of the paper. One aspect of the answer to the question of how individuals can get a grip on the multiplicity of affordances available already becomes clear from looking at the structure of the landscape of affordances.

1.2.1 The structure of the landscape of affordances

The concept of a "*landscape* of affordances" aims to capture the interrelatedness of the available affordances. Affordances are not encountered as a set of separate possibilities for action, but rather as a nested structure of interrelated affordances.[3] In the case of the form of life of enculturated human beings, this structure can be very complex. It is only against the background of socio-cultural

---

[3] This is a remark that concerns the structure of the ecological niche and not our phenomenology. Phenomenologically, the structure of our experience of solicitations resembles more that of a field with some solicitations standing out and with a horizon. Moreover, this is not to deny the fact that our lives proceed along paths or trails, as Ingold (2011, p. 147) rightfully stresses, "always on the way from one place to another."

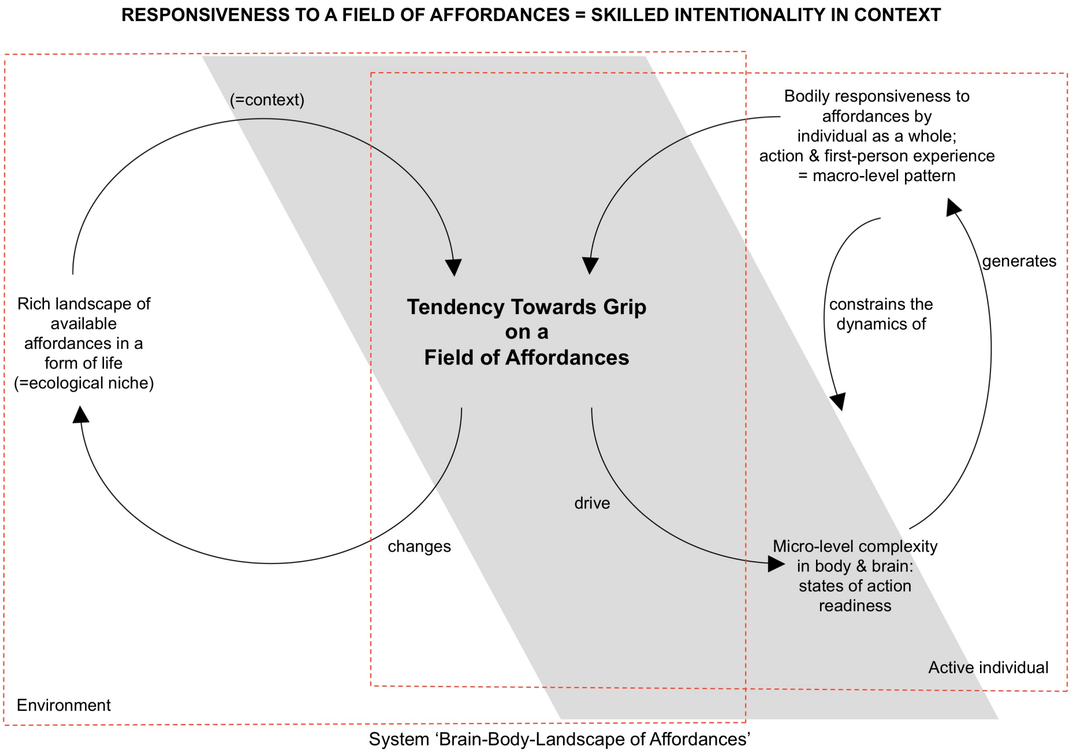


Figure 1.1: Sketch of the conceptual framework to be refined. Through skilled intentionality one gets a grip on a field of affordances (Rietveld, 2013) (inspired by Chemero, 2003, 2009; Dreyfus, 2007; Thompson, 2007, 2011; Tschacher & Haken, 2007; Rietveld, 2008a, 2008b, 2012a, 2012b).

practices, places and institutions that the affordances here in my office are intelligible. The affordances of places (libraries, restaurants, etc.) typically constrain behavior over a longer timescale, while the affordances of objects nested in such a place, say the door to the library's reading room, typically constrain behavior on a shorter timescale.[4] Such place-affordances (the affordances of say, university libraries, railway stations, supermarkets, swimming pools or restaurants) are the contexts in which many of our activities unfold (Kiverstein & Rietveld, 2012; cf. Heft, 2001). Which affordances are relevant depends on the "behavior setting" (Barker, 1968; Heft, 2001): the possibility of calling a waiter is relevant in a restaurant but not when we are in a supermarket. Being in a restaurant constrains or pre-structures which affordances are relevant to me. In order to be

---

[4] We do not wish to claim that the landscape has a clear hierarchical structure. Rather, the structure would be more like a heterarchy. That is to say, there is no strict demarcation of levels within the nested structure, although when focusing on a specific event, such as dining in a restaurant, place-affordances can be discerned.

responsive to the appropriate affordances of a situation (e.g., calling out a waiter in a restaurant), one needs to be well attuned to the current context (one needs to have the ability to deal and be ready to deal with restaurants and waiters). In sum, we suggest that responsiveness to a place-affordance, which is a nest of affordances, generates an action-readiness that makes the individual selectively open to the landscape of affordances. As such this responsiveness pre-structures the relevance of locally available affordances in a way that allows the individual to have a grip on the rich landscape of affordances in which she is situated.

The nestedness of the landscape of affordances thus helps the organism to gain a grip on multiple relevant affordances simultaneously. The challenge for the organism is to, in a particular situation, be selectively open to only the relevant affordances. In the remainder of this paper we seek to find out how theoretical neuroscience and dynamical cognitive neuroscience contribute to understanding such self-organized relevance sensitivity.

1.3 Self-organization

One of the developments relevant for an understanding of the mechanisms that contribute to selective affordance-responsiveness is an improved understanding of self-organizing systems. Especially, we are interested in self-organizing systems that are able to actively influence their interactions with the environment in order to adapt to and induce environmental changes, i.e., so called homeokinetic or self-serving systems (Iberall, 1977; Turvey & Carello, 2012). The theory of self-organization is particularly suitable for the framework of affordance-responsiveness developed here, because in both of these theoretical frameworks, it is the reduction of a tension or gradient that is the central motivation for an action: it is the environment that is the driving force for an action for an organism in a particular situation. We will first present the familiar Bénard effect as an example of how self-organizing patterns can be functional with respect to their environment and subsequently describe how the theory of self-organization can improve our understanding of affordance-responsiveness.

Self-organizing systems are initially disordered systems where global order can arise under the influence of the system's own dynamics. This is typically the case when a control parameter reaches a critical value upon which new forms of organization become possible for the system. Within the self-organizing range, the behavior of the system is low dimensional, i.e., it can be quantified by a small amount of order parameters that describe the macroscopic patterns in the system (See **Table 1.2**). Classical examples from the literature stem from diverse fields such as treatments of the Bénard cell in non-equilibrium fluid dynamics (Bénard, 1900; Bishop, 2008), the laser in optics (Haken, 2004) and coordination dynamics in cognitive science (Haken, Kelso, & Bunz, 1985).

State space: The space defined by the set of all possible states a system could ever be in.

Trajectory (path): A set of positions in the state space through which the system might pass successively. The behavior of the system is often described by trajectories through the state space.

Attractor: A point of state space to which the system will tend when in the surrounding region.

Topology (attractor landscape): The layout of attractors in the state space.

Control parameter: Some parameter of a system whose continuous quantitative change leads to a non-continuous, qualitative change in the attractor landscape.

Order parameter: Some parameter of a system that summarizes the behavior of the system's components.

Circular causality: The mutually constraining relationship between the microscopic and macroscopic elements of a complex system: the order parameters emerge out of the microscopic dynamics, while the order parameters themselves constrain or enslave the microscopic dynamics.

Second circularity: The mutually constraining relationship between one or more control parameters in the environment and a self-organizing system. The system self-organizes in order to reduce the control parameter(s) that gives(s) rise to its self-organization.

(Central) pattern generator: A dynamical system producing rhythmic patterned activity potentially modulated by feedback mechanisms.

Metastability: A property of coupled dynamical systems in which over time the system's tendency to integrate and segregate coexist.

Table 1.2: Terminology of complex and dynamical systems These are all standard definitions, in this case obtained from Chemero (2009); Kelso (2012); Rabinovich et al. (2008); Tschacher (2010).

1.3.1 Rayleigh–Bénard convection

The Rayleigh–Bénard effect is empirically, theoretically and philosophically the most well studied non-linear self-organizing system. The phenomenon occurs when a layer of fluid is heated from below. Cold water is denser (hence heavier) than warm water, so the temperature difference creates a buoyancy force. When the temperature difference is small, the viscosity of the fluid counteracts the buoyancy force and the system will dissipate energy through heat conduction. When the temperature gradient passes a critical value, the buoyancy force overcomes the viscosity (more potential energy is brought in the system than can be dissipated through heat conduction) and the system becomes globally unstable. This leads to convection patterns in the shape of parallel cylinders (so called convection or Bénard rolls).

In the formalization of the Bénard effect, the temperature difference between the top and the bottom of the fluid is considered a control parameter. The macroscopic state of the system (conduction or convection) is a function of the control parameter. Furthermore, in the self-organizing regime, the system can be described and determined by only a few variables, the so-called order parameters. The relation between the order parameters and the microscopic components (the single molecules of the liquid, e.g., water molecules) is a peculiar one: the order parameters constrain the trajectories of the parts, but the parts also generate the order parameters. The relationship between parts (the microscopic) and whole (the macroscopic) is one of mutual constraints or, to use Tschacher and Haken's philosophically somewhat problematic term, *circular causality* (Tschacher & Haken, 2007).

1.3.2 Gradient reduction and second circularity

How can the theory of self-organization help us to understand the mechanisms of the tendency toward an optimal grip in human beings? There is a second fact about self-organization in the Bénard system. The self-organization has an impact on the environment as well. The self-organization reduces the very temperature gradient that gives rise to it: it is the temperature difference that enables the convection, but the convection reduces the temperature difference. It is due to this so called second *circularity*, that self-organized patterns are functional with respect to their environment, that is to say: the patterns are *geared toward the* *reduction of the environmental gradients*[5] on the system. Crucially, the *function* of self-organized pattern formation, according to Tschacher and Haken (2007), is to adapt to environmental constraints and realize dissipation of the gradients.

It is these two circularities that we find in affordance responsiveness as well. On the one hand, solicitations move the organism in a particular direction; on the other hand leads the responsiveness to the solicitation to a reorganization of the field of affordances, which makes new solicitations stand out. We therefore propose to think of relevant[6] affordances as gradients that drive the dynamics of the system and in return are consumed by it.

There is, however, an important difference between a Bénard system and a system like the brain-body-environment system: in the Bénard effect and most other standard examples of self-organization, there is only one control parameter working on the system. For our purpose of understanding the mechanisms of optimal grip in the case of human beings, it is important to consider the case of multiple control parameters, because generally there are multiple relevant affordances in any particular situation of an individual.[7]

1.3.3 Self-organization and living systems

There is another significant dissimilarity between systems like the Bénard system and systems like the affordance-responsive organism. In the case of non-living systems, as in the Bénard system, the self-organizing pattern disappears if the external control parameter decreases below a threshold. For example, if the temperature difference reaches below the critical value, the organized patterns disappear. Living systems have to be able to actively interact with the gradients that affect their self-organization. One could then say that the gradient is not *given by*, but *obtained from* the environment (Iberall, 1977; Turvey & Carello, 2012). In the first case, systems are served by the environment, while in the second case,

---

[5] The notion of gradient has a clear physical interpretation in the case of the Bénard effect: it is the difference in temperature between the top and the bottom of the layer of fluid. In the case of the coordination dynamics of locomotion, the gradient is for instance the speed of the treadmill to which the animal adapts its gait. Tschacher and Haken (2007) give an example of a psychological gradient guiding an action: in the context of a letter that one has to mail, the letter-affording-delivery stands out as a gradient to be reduced.

[6] Tschacher and Haken (2007) do not make the distinction between solicitations and affordances. Their use of the word "affordances" applies to gradients that actually drive the system (i.e., what we call solicitations).

[7] An important open question is that of optimality: on some interpretations of self-organization (Schneider & Kay, 1994), the pattern that arises is always the one that most efficiently (i.e., in the least amount of time) dissipates the gradient. As Haken and Tschacher (Haken & Tschacher, 2010) point out, it is not clear that such an optimality principle for self-organizing systems in general is feasible.

systems are *self-serving* or homeokinetic.[8] These latter systems can internally generate forces to counteract the effect of physical gradients on the system, and move through their material environment to avoid harmful gradients and find new ones [this is what Turvey and Carello (2012, p. 11) call "proto-foraging" behavior]. Crucially, through this capacity, the system is able to (within limits) influence the gradients that affect it and hence maintain its own self-organization (Kugler & Turvey, 1988; Turvey & Carello, 2012). In the hypothetical case of a *living* Bénard cell, this would amount to a layer of fluid being able to heat or cool itself, or to move through a temperature landscape in the environment in order to regulate its self-organizing patterns.

What is interesting about Haken and Tschacher (Haken & Tschacher, 2010) proposal is the conceptual link between gradients and affordances. They do emphasize that the reduction of gradients can also occur when more gradients work on a system, but in their (2007) account, the nature of these gradients and their structure remains undeveloped. The perspective we have sketched advances Tschacher and Haken's account of affordances in three ways. First, we distinguish conceptually between affordances and solicitations (Rietveld, 2008a; cf. Rietveld, 2008b; Withagen et al., 2012). Second, we show that each affordance is embedded in a landscape of affordances of a given form of life, which includes socio-cultural practices in our human form of life. The embeddedness in this landscape is crucial for adequate anticipation of the organism in its environment. It is only when we are attuned to the specific context "including place-affordances" that we can adequately be responsive to relevant solicitations that are in line with our concerns. Third, at the level of the individual as a whole we connect the reduction of gradients with the tendency toward an optimal grip on a concrete situation.

Our formulation of affordance-responsiveness in terms of self-organization does not yet address the problem of context-sensitive *selective* openness to affordances, which, as we have suggested in the introduction and earlier work (Kiverstein & Rietveld, 2012), should be the central topic of Radical Embodied Cognitive Neuroscience. The theories of self-organization and synergetics (Haken, 1983) provide the framework in which to investigate this important problem. In the upcoming sections of this paper we explore how a complex system like the brain can be selectively sensitive to only *some* environmental gradients/affordances.

---

[8] Iberall writes: "[Self-serving systems] can explore its surround to acquire the necessary potentials at its boundary that serve as sources of free-energy for its own internal and externalized processes. In this case internal processes convert internal energy into a useful form of work that can change momentum and move the system to a favorable location" (1977, p. 177).

1.4 Anticipation and selective openness

In recent years, there has been growing interest in the application of ideas from statistical physics, machine learning and complex and dynamical systems theory to the brain (see for instance Freeman, 1987, 2000; Friston, Kilner, & Harrison, 2006; Tognoli & Kelso, 2014). What these approaches have in common is their appreciation of the brain as an intrinsically active and unstable self-organizing system. In part thanks to these authors, progress has been made in how the self-organization of the brain can be functional with respect to the larger brain-body-environment dynamics (see also Freeman, 2000; Dreyfus, 2007). We think that this perspective (neurodynamics embedded in brain-body-environment dynamics) is the natural starting point to develop a Radical Embodied Cognitive Neuroscience.

One promising proposal to couple brain, body and environment is Karl Friston's FEP (Friston, 2010).[9] According to the FEP, any self-organizing system that remains within physiological bounds in its interactions with a changing environment (and hence resist a natural tendency to disorder), can only frequent a limited amount of physical states. This can be given a mathematical interpretation in the sense that the probability distribution of the organism's states must have low entropy (i.e., there is a high probability that a system is in one of a relatively small number of states). This long term imperative to constrain the entropy of its states translates into a short term imperative to suppress surprisal[10] (see **Table 1.3**). Importantly, surprisal can not be suppressed directly, since it depends on the expected range of states over time. The information theoretic quantity of free-energy (not to be confused with the homologous concept from thermodynamics)[11] is an upper bound on surprisal such that when an organism minimizes free-energy, it is implicitly minimizing surprisal (Friston, 2011).

In the active inference formulation (Friston, 2010; Friston et al., 2013) of the FEP, free-energy can be minimized on short time scales by making the environment conform to the internal dynamics ("action") or by making the internal dynamics conform to the environmental dynamics ("perception"). There is an important similarity between Tschacher and Haken's framework of self-organization and Friston's FEP: what they call circular causality and second circularity map onto what Friston calls "perception" and "action", respectively. It is through these two circularities that organism and environment are coupled.

---

[9] What follows is a treatment of the theory of the FEP. For mathematical details, see Friston et al. (2006); Friston (2012a).

[10] Because under ergodic assumptions, entropy is equal to the average of self-information (surprisal), see Friston et al. (2009) for mathematical details.

[11] The latter has a clear physical definition in terms of the amount of energy available in a system that is convertible to work. The former is a quantity from information theory, which is an upper bound on surprisal. As such, information theoretic free-energy has nothing to do with energy in the ordinary sense of the word.

Surprisal: A measure for the unexpectedness of an event expressed in terms of the negative log-probability of the event outcome.

Free-energy: An information theoretic measure that is an upper bound on the surprisal of some data, given a generative model.

Prediction error: The difference between anticipated and actual sensory input. Under simplifying assumptions, free-energy equals the sum of prediction errors.

Table 1.3: Information theory and the anticipating brain. Standard definitions taken from Friston (2010).

The FEP in itself makes no claims about the mechanisms underlying free-energy minimization. It is supposed to be a necessary requirement for any adaptive self-organizing system that is able to resist the tendency to disorder. When it comes to organisms with developed nervous systems, the FEP offers a rich and sophisticated set of tools in order to gain a better understanding of how free-energy can be minimized. Given some simplifying assumptions (cf. Marreiros, Kiebel, Daunizeau, Harrison, & Friston, 2009) the brain dynamics can be modeled using variational Bayesian methods and hierarchical predictive-coding. However, to avoid misunderstandings, it is important to distinguish between the imperative (i.e., minimizing free-energy) and the mechanisms by which the organism obeys that imperative. As Friston himself notes: “The Bayesian brain and predictive-coding are [...] seen as a consequence of [...] this fundamental imperative [of free-energy minimization.]” (Friston, 2013a, pp. 212-213). Free-energy minimization is thus the primary notion and we wish to foreground that, rather than the Bayesian and the predictive-coding framework.[12]

The FEP implies a deep connection between the dynamics of the brain-body-environment system and the neurodynamics. What is crucial, for the organism, is that it anticipates the kind of interactions with the environment that lead to an adequate outcome (such as having food, or avoiding a passing car). The function of the generative model is therefore not to provide the agent with a representation of the dynamical structure of the environment *per se*, but rather to steer its interactions with its environment in such a way that a *robust* brain-body-environment system is maintained. The internal dynamics, Friston’s generative model, can not be understood apart from its functioning within the integrated brain-body-econiche system.

---

[12] Within the context of the FEP, much attention in the literature has been given to how efficient information processing is possible (in the form of predictive-coding and approximate Bayesian inference), however, much less attention has been given to what structures in the environment the anticipating organism is responsive to. In this paper we are concerned with the latter question.

To illustrate this point, note that Friston himself states, somewhat provocatively, that: “each [...] agent embodies an optimal model[13] of its econiche” (Friston, 2011). Furthermore, Friston states that:

> [A]n agent does not *have* a model of its world–it *is* a model. In other words, the form, structure, and states of our embodied brains do not contain a model of the sensorium–they *are* that model. [...] But what does this mean practically? It means that every aspect of our brain can be predicted from our environment (Friston, 2013a, p. 213)

For Friston, the niche implies the structure of the organism. Now, for our argument, we do not need to subscribe to this last claim in the fullest sense, but it shows the radical potential of the FEP.

In general we think the FEP is a step forward in understanding the relation between environmental dynamics and neurodynamics. It is an attractive framework because we think it is able to formalize the tendency toward an optimal grip in terms of the dynamical coupling between brain dynamics and the dynamics of the whole brain-body-environment system, or more specific: of the whole system “brain-body-landscape of affordances”. Within the framework of the FEP the tendency toward an optimal grip could be seen as a consequence of the continuous minimization of free-energy through perception and action at the level of the organism as a whole: *the attunement of the internal dynamics and external dynamics.*

However, we worry that along with the welcome mathematical sophistication comes a vocabulary that is mathematically convenient, but philosophically problematic (Anderson & Chemero, 2013). For instance, within philosophy and cognitive science the notion of “inference” is traditionally understood in terms of arriving at a propositional statement based on some premises or observations. Within the free-energy framework, the notion of “inference” is much more minimal and does not involve any propositions: any dynamical system A coupled with another B can be said to “infer” the “hidden cause” of its “input” (the dynamics of B) when it reliably covaries with the dynamics of B and it is robust to the noise inherent in the coupling. [For a presentation of this minimal notion

---

[13] Although the FEP uses the word “model”, we think that it is used in a way that makes it sufficiently compatible with radical embodiment. What radical embodiment is against, is the idea that an agent has an internal model of the world, which, through some inference process, provides the agent with a representation of the world on which it consequently can decide what to do. This is not what the FEP entails.

of inference, see Friston, 2012a, 2013b]. This is important, because it suggests that the apparent tension between radical embodiment and the FEP is at least to some extent terminological.[14]

To summarize, the FEP dictates that in order to maintain a robust brain-body-environment system, an organism can and needs to continuously minimize the prediction error or discrepancy (formalized in terms of free-energy) between its internal dynamics and the dynamics of the larger system. The organism does not need to have a model of its niche, but rather the claim is that the structure of the niche is reflected in the structure of the skilled embodied organism. We will argue that the internal dynamics should be understood in terms of affordance-related action-readiness patterns. The notion of an econiche is not developed any further in Friston's work up to now, but we will come back to the relation between an organism's niche (made up of a landscape of affordances) and the internal dynamics in **Section 1.6**.

So far we have focused on integrating our theoretical framework of skilled intentionality with the theoretical framework of the FEP. The integration of these two frameworks now places us in a position to look at the neurodynamics of selective affordance-responsiveness under the FEP. It is here that the theory of self-organization, introduced in the previous section of this paper becomes important again

1.5 The neurodynamics of selective openness

In this section we will present a neurodynamical approach that is able to account for selective-responsiveness to affordances within the adjusted framework of the FEP. Within the free-energy framework, selective responsiveness is brought about by pattern generators that make both sensory (exteroceptive) and motor (proprioceptive) predictions (Friston, Shiner, et al., 2012).[15] Pattern generators are well known through the work of Randall Beer on robot locomotion (Beer & Chiel, 1993). They are systems that are capable of producing rhythmic or sequential patterns and can be modulated by sensory feedback. Beer uses coupled pattern generators with sensory feedback to build distributed control circuits for robot locomotion. The dynamics of a pattern generator is modulated and constrained by both its sensory feedback and the dynamics of the other pattern generators.

---

[14] The radical response would be to question the added explanatory value of the notion of inference over and above the dynamical explanation (Chemero, 2009). We lack the space here to retranslate the FEP in non-propositional, dynamical terms, but we think that this is possible. For the moment, it is important to emphasize that notions such as "inference", "belief", and "expectation" all have a different meaning within computational neuroscience and philosophy.

[15] In fact, Friston uses the word "affordance" to designate the activation patterns that guide affordance responsiveness. This is not in line with how the term is traditionally used in ecological psychology (Gibson, 1979) and philosophy (Chemero, 2003) and is bound to lead to confusion. We will use "action-readiness pattern" to designate what Friston calls "affordance."

Kiebel, Von Kriegstein, Daunizeau, and Friston (2009) show that by coupling pattern generators evolving at different timescales, one can create a dynamical system (a generative model in the sense introduced in the last section) that is capable of swiftly interacting with a complex dynamical environment. The pattern generator evolving at longer timescales serves as a control parameter that shapes the attractor at which the lower-level dynamics unfold. The specific kinds of pattern generators they use are so called stable heteroclinic channels (Rabinovich et al., 2008). These are defined as a sequence of metastable (saddle) points with transients in between.[16] When these stable heteroclinic channels are coupled in a temporal hierarchy, the ensuing dynamics never reaches a fixed stable point, but continuously follows a trajectory through state space (Kiebel et al., 2009). This trajectory is continuously modulated through sensory feedback (prediction errors). Some prediction errors can be accommodated for on the lower level, leaving the slower-evolving patterns intact (for instance when synchronizing to an external rhythm), while other prediction errors, can induce or destroy the pattern generators at a longer timescales as well (such as when the beat of the music changes dramatically).

This is important for understanding how the selective openness helps to make, in the particular situation, the distinction between the relevant affordance(s) and other affordances; between the one(s) to be responded to here and now and the ones that leave the organism cold. The generation of an adequate action-readiness rests upon precise sensory feedback that feeds into a dynamical system (generative model) that is shaped by the organism's previous interactions with the environment. The system will settle on a pattern that explains away most of the prediction error (i.e., the system tends toward a particular attractor). On slower time-scales this amounts to "action selection", while on the faster timescales the action is specified: prediction errors influence the attractors that make more specific sensorimotor predictions ("action specification"). Both action selection and action specification depend on sensitivity to small disturbances that is, deviations from anticipations generated by pattern generators (Cisek, 2007; Cisek & Kalaska, 2010).

The fact that stable heteroclinic channels implement metastable attractor dynamics is crucial for understanding the flexibility of selective openness to affordances. Kelso (2012) describes metastability as the outcome of two competing tendencies: the tendency of the components to couple together and the tendency to express their independent behavior. In this metastable regime, the system is poised at the edge of instability, a kind of dynamic stability that allows the system to maintain "a balance in the readiness of the system to transit between multiple attractors" (Davids et al., 2012, p. 119). While being skillfully engaged

---

[16] An intuitive example of a stable heteroclinic channel would be a pub-crawl. One visits a sequence of bars (the metastable saddle points), while walking from one to another in between (the transients). The sequence might be fixed, but the timing for when to move to a new bar is generally left to the specific circumstances.

with a specific task, it is important that we can be affected by affordances on the horizon of our field and rapidly switch to another kind of adequate activity when something in the environment changes. Metastable dynamics are important for understanding the brain, because metastability is a prerequisite for a system to be able to effortlessly switch between different patterns. We will see that metastability plays an important role as well in the brain-body-environment dynamics of skilled agents, in **Section 1.7**.

In Friston's picture, the elicitation of an action-readiness-pattern triggers a cascade of spatiotemporal dynamics in the brain modulated by sensory input that aids anticipation on the interactions with the environment. In ballroom dancing for instance, the first measures of music will afford either dancing tango or waltz. The elicitation of the tango-dancing-pattern will trigger an attractor-manifold that governs the sensorimotor coordination between me, my dance partner and the music: this action-readiness pattern will make certain action possibilities solicit more to me than others. On a more fine-grained level, small cues by the dance partner and subtle variations in the rhythm in the music further specify my action-readiness. Only if I am well attuned to the context (the situation) and thus metastably poised for several relevant activities I could do next, can small cues in the environment lead to very different positions in state space and hence to flexible responsiveness to (very) different solicitations. That is, only when I am able to rapidly accommodate the small deviations from my anticipations (in Friston's terms: the ability to explain away prediction errors through perception and action) can I engage skillfully with a complex environment.

Within our adjusted version of the FEP, a solicitation is a gradient/prediction error that, through action, can be resolved by a change in the brain-body-environment system. These gradients are the result of the individual's selective openness to the available affordances which is the result of dynamical patterns evolving at multiple time-scales. The dynamics unfolding over long timescales act as control parameters or constraints for dynamics unfolding over shorter timescales. Crucially, when the dynamical system (generative model) and the environmental dynamics are well attuned to each other, the solicitations/gradients/prediction errors that stand out as to-be-responded-to are the ones that lead toward an optimal grip on the environment.

An open question that remains is the following: what does it mean to say, under the FEP, that the organism and the environment are well attuned to each other? In other words, what aspects of the environment must the generative model be reflecting for the organism to interact adequately with its environment? We will address these questions in the next section.

1.6 Situating the anticipating brain

Radical Embodiment emphasizes the non-decomposability of the brain-body-environment system, which implies that the neural dynamics can only be studied while taking into account the larger brain-body-environment dynamics (Chemero, 2009). When focusing on one element of these dynamics, such as the brain, one can model the rest of the dynamics as control parameters (Friston, 2000a). This allows for several perspectives on essentially the same dynamics: the state variables of the brain-body-environment system can be control parameters for the brain. From this perspective, it is possible to focus on the dynamics of the brain:[17] in this case, the body and the environment are described as control parameters (prediction errors) that are changing themselves. Given that the brain is situated within a robust brain-body-environment system, one can derive constraints on how the brain is coupled to the wider system. Following this analysis of the dynamical coupling, one ends up with the perspective of the FEP.

If aspects of our brain can be predicted from our environment, we need to understand which aspects of the environment are being reflected in brain dynamics. The fundamental idea of the FEP is that by being equipped with a generative model that reflects the hierarchical and temporal organization of the changing environment, organisms are able to remain attuned with the dynamics of the environment. This invites the question how the landscape of affordances, introduced in the first section of this paper, and the generative model/the organism are related to each other.

At several places Friston states that the agent is inferring the causal structure of the environment (e.g., Friston, 2011). However, it is important to qualify this in several respects. First, above, we have interpreted Friston's notion of inference in a non-propositional way fully within the domain of dynamical systems. Second, the agent is not modeling the causal structure of the environment *per se*, but rather those aspects of the environment that are important within its specific niche. We think that what is "inferred" in active inference, as we have noted above, are not objects or properties of objects, but rather anticipatory patterns that specify a solicitation. A pattern on which the system settles does not *represent*, say, a carrot, the smell of a carrot, or what to do with a carrot, but rather, the attractor state is directly coupled to the affordance of the carrot here and now (Freeman, 2000; Dreyfus, 2007): at no point in skillful action is the organism inferring the current causal state of the environment, and *on top of that* figuring out what change in the causal structure will lead to a more favorable outcome. Rather, the gradients/prediction errors themselves trigger the right anticipatory

---

[17] Note that this pertains to the domain of coupled dynamical systems. Given the centrality of the brain-body-environment system as a whole, we do not think the possibility of constructing such a perspective from the brain justifies epistemic internalism.

pattern that makes the right affordance stand out and that minimizes free-energy or, in more phenomenological terms, leads to an optimal grip on the organism's environment.

Inspired by Gibson Gibson (1979) we have, as mentioned in the introduction, suggested that we can understand the ecological niche as a landscape of affordances (Kiverstein & Rietveld, 2012). Armed with our understanding of the richness of the landscape of affordances available in our form of life (as developed in the first part of the paper), we argue that what the embodied organism is "modeling" or reflecting in a particular situation, is not so much the causal structure of the environment *per se*, but rather the dynamic nested structure of the field of affordances. We do not think this is in contradiction with the FEP but rather a natural consequence of combining active inference (action and perception jointly reducing gradients/prediction errors) and the need for the organism to be governing its interactions with the environment.

This contextualization of the anticipating brain is important for two reasons. First, it makes clear that the FEP really calls for an integrative approach for understanding the mutual attunement of the brain and the other components of the whole brain-body-environment system. The deep correspondence between the dynamics in the environment and the neurodynamics implies that we can learn something about the brain by investigating the structure of the econiche, i.e., of the landscape of affordances.

Second, it provides a new understanding of the tendency toward an optimal grip, which is a central notion in phenomenology, as the concernful skilled agent's tendency to reduce his or her dis-attunement to the environmental dynamics. In particular, it provides an understanding of how the relevance of affordances is selectively brought: the relevance of an affordance (an attribute of the brain-body-environment system) is in part brought about by aspects of the environment triggering patterns that shape the skillful agent's action-readiness for interacting with its environment. We think that the field of affordances both captures an important aspect of the phenomenology of skilled intentionality, and can inform theoretical neuroscientists about what it is the self-organizing brain is responsive to (i.e., what external control parameters influence the self-organization of the brain). Skilled intentionality should be of particular interest to those who work on the implications of the free-energy principle, because it is the kind of intentionality manifested when we act as "surprisallessly" as possible: when we are in familiar environments and can act relatively unreflectively and effortlessly.

1.7 Toward a Radical Embodied Cognitive Neuroscience

In the previous sections, we have presented an integrative framework for studying skilled intentionality. In this section we will illustrate the plausibility of our framework by presenting work on metastability in the system "brain-body-landscape of affordances" dynamics of skilled sportsmen, and empirical research on the impact of DBS on affordance responsiveness of OCD patients.

1.7.1 Metastability and optimal grip

Above we have seen that metastable dynamics are an important characteristic of neurodynamics, because it allows for context-sensitive selective openness and flexible switching between activities. An interesting property of metastable dynamics in the brain, like the stable heteroclinic channels described in **Section 1.5** for example, is the possibility to be both robust to perturbations and flexible.[18] The dynamics of the coupled patterns generators can be described as visiting a succession of unstable fixed points in an abstract state space (Tsuda, 2001; Rabinovich et al., 2008). The itinerant dynamics can be observed at different time-scales or at different levels of the hierarchy. One can see how such a system can be both robust and flexible: on the one hand do slower-evolving dynamics constrain the faster-evolving dynamics, on the other hand, because of the metastable character of the slower dynamics, some perturbations (e.g., as a result of gradients/prediction errors) can easily and swiftly change the slower dynamics and make it shift to a new pattern that better fits with the multiplicity of affordances currently encountered.

Importantly, metastable dynamics in the brain-body-environment system as a whole provide an important paradigm for understanding movement pattern variability in ecological situations. For example, Hristovski, Davids, Araújo, and Button (2006); Hristovski, Davids, and Araujo (2009) investigated how boxers' striking patterns differed when manipulating the distance to a boxing bag. At great distances, they observed a "jab" movement, while at short distances, they observed "hooks" and "uppercuts." At a critical distance of 0.6 (the distance to the punch bag scaled by the arm length), they found an optimal metastable performance region where a varied and creative range of movement patterns occurred: a region in which the boxers "could flexibly switch between any of the boxing action modes" (Chow, Davids, Hristovski, Araújo, & Passos, 2011, p. 197). So, at different scaled-body distances, the boxing bag solicited different punches, but at the optimal metastable distance, the boxing bag solicited a wide variety of punches. Here something occurs that might be called a *Hypergrip on*

---

[18] This contrasts with phase locked dynamics. Phase locked dynamics are generally robust to small perturbations as well, but lack the flexibility of metastable dynamics.

the field of affordances (Rietveld, 2013). For an expert boxer the zone of optimal metastable distance will solicit moving toward, because this zone offers a wide range of action opportunities and the possibility to flexibly switch between them in line with what the dynamically changing environment demands or solicits.

Anticipation is an important aspect of the phenomenon of Hypergrip on the field of affordances. This is best illustrated by means of an example from a different field of expertise. In ice-climbing, the metastable regime is one where the expert climber can use different movement patterns to obtain the same result (Seifert et al., 2014). Moreover, a skilled climber is anticipating the affordances ahead; she does not just get a grip on the next hold in climbing, say, but also anticipates that she needs to be able to move on after that. So, the question of relevance sensitivity is not just about grasping the next hold, but rather about which of the available holds afford obtaining a grip on the *whole climbing route ahead*. One can see again that in such a metastable state, one is flexibly able to switch between different movement regimes and better fit to adapt to the specific details of the environmental aspects.

These studies suggests that, at least in some domains of skilled action, we can formalize the tendency toward an optimal grip in terms of the occurrence of metastable movement patterns. More precisely, we can understand the tendency toward an optimal grip as the tendency toward an *optimal metastable attunement* to the dynamics of the environment. This optimal readiness to switch between behavioral patterns is both functional with respect to the demands of the environment and the needs of the organism.

Further empirical research on optimal metastable performance regions in ecological psychology will thus be able to illuminate the phenomenon of the tendency toward an optimal grip and the selective openness to relevant affordances. It will be particularly interesting to see what agents will do in situations in which there is not a specific task given, or when they are allowed to switch spontaneously between different ways to solve a task, just like in everyday life.

Moreover, the phenomena of flexible switching and Hypergrip on the field of affordances on the horizon touch upon one of the most important open questions in cognitive science, the frame problem (Wheeler, 2008; Rietveld, 2012b). Skilled intentionality treats context as just more affordances–a landscape of affordances available in an ecological niche–and avoids the frame problem by starting from the phenomenon of maintaining grip on multiple affordances simultaneously.

How can the neurodynamics involved in selective openness support an optimal grip on the whole field of affordances including possibilities for action on the horizon? In order to answer this question, we need to understand how the self-organized metastability of the brain-body-environment system interacts with the self-organized metastability of the brain. To advance, it is important to develop neuroscientific research methods that are able to complement the work done on boxing and climbing in an actual ecological setting. What is the difference in neurodynamics in the optimal metastable region as compared to the other regions of performance? In the next section, we present recent research by our group on the impact of DBS on affordance responsiveness as an example of such a complementary approach that also takes phenomenology seriously.

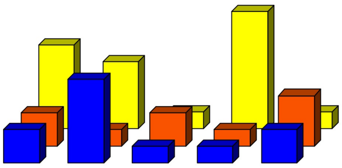
1.7.2 Mood disorders and relevance-sensitivity

Our understanding of relevance sensitivity as the self-organized coordination of action-readiness patterns is closely related to Frijda's theory of emotions (Frijda, 1986, 2007). According to his theory, the key aspect of an emotion is a state of action-readiness for changing an aspect of the self-object relationshi p.

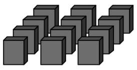
Moods are action-readiness patterns that persist for longer periods of time and typically have a relatively global character: a mood is reflected in the structure of field of affordances as a whole (see **Figure 1.2A**). Similarly, we can understand mood disorders as disorders that distort the field of affordances: in the case of depression, for example, the field of affordances is rather flat, nothing stands out as attractive or soliciting anymore (see **Figure 1.2B**). A recent qualitative study investigates the effects of DBS on the phenomenology of patients suffering from treatment resistant OCD (De Haan et al., 2013). OCD can be characterized by the presence of anxiety-provoking thoughts, typically followed by ritualistic behaviors (compulsions) to relieve the anxiety. In extreme cases of OCD, the patient's field of affordances is narrowed down to just the immediate solicitation of what has to be done here and now without the possibility of flexibly switching to a new readiness for other behaviors (see **Figure 1.2C**).

DBS treatment consists of permanently implanted electrodes that deliver electrical pulses to a target brain region. DBS of the nucleus accumbens shows promising results as treatment for OCD patients (Denys et al., 2010). It is hypothesized that, rather than merely having inhibitory or excitatory effects on the target area, DBS restores intrinsic brain network dynamics (Figee et al., 2013). In particular, the authors show that DBS treatment normalizes the activity of the nucleus accumbens and restores the intrinsic frontostriatal network dynamics. These frontostriatal circuits are known to be important for switching between different actions (Ridderinkhof, Forstmann, Wylie, Burle, & van den Wildenberg, 2011). Furthermore, it was found that the frontostriatal connectivity changes strongly correlated with OCD symptom improvement (Figee et al., 2013).

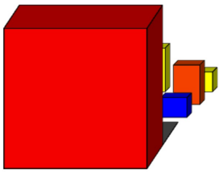
From phenomenological interviews with these OCD-patients, it becomes clear that, by treatment with DBS combined with cognitive behavioral therapy, these patients report a general change in engagement with the world pertaining to perception, reflection, mood, interests and social interaction (Rietveld et al., 2013). These impressive phenomenological changes can be understood as changes in the responsiveness to the field of affordances along three dimensions: the "width" (the scope of affordances engaged with), the "depth" (the temporal horizon), and the "height" (the relevance) of the field of affordances (see **Figure 1.2**) (De Haan et al., 2013).



1a. Normal



1b. Depression



1c. OCD

Figure 1.2: Sketch of different fields of relevant affordances. (A) A normally structured, differentiated field of affordances. (B) The field of affordances of a depressed person. (C) The field of affordances of an OCD-patient. Creative Commons license applies (De Haan et al., 2013).

Above we have argued that affordance-responsiveness corresponds to heavily interacting neurodynamics at different temporal scales. DBS can directly perturb these neurodynamics and hence directly influence the general capability of affordance-responsiveness (De Haan et al., 2013). The observed change in affordance-responsiveness opens up the possibility to develop neurodynamical models of OCD and psychiatric disorders more generally, based on Friston's ideas on the anticipating brain. Friston's model of addiction (2012b) highlights the importance of metastable, itinerant dynamics for modeling adaptive and pathological behavior. In personal communication (June 15th, 2012) Friston has suggested that OCD could also be modeled along the same lines. We are currently working on a research project that uses our framework and an updated version of Friston's (2012b) model for addiction, to generate testable hypotheses about the neural mechanisms that could underlie the breakdown of normal affordance-responsiveness in OCD-patients and its recovery through DBS treatment.

1.8 Conclusion

In this paper, we have investigated the phenomenon of skilled intentionality from the perspective of the self-organization of both the brain-body-environment system and the brain. Previously, we have characterized skilled intentionality as the organism's tendency toward an optimal grip on a field of relevant affordances. In this paper we have investigated the mechanisms that underlie the self-organized selective openness to available affordance and the organism's tendency toward an optimal grip. We have integrated different perspectives on this phenomenon: the philosophy of skilled of intentionality, Kelso's and Tschacher and Haken's ideas on self-organization, Friston's theory of the anticipating brain, and work on metastable dynamics. What these four perspectives on the affordance-responsive active individual have in common is the idea that a an organism self-organizes by reducing a disequilibrium in the brain-body-environment system. On the different levels of analysis, this disequilibrium can be called a solicitation, a gradient or a prediction error, or a dis-attunedness of internal dynamics and environmental dynamics.

Our integrated framework moves beyond the traditional Gibsonian conception of affordances, because it highlights that the animal cares about certain things and needs to be selectively open to the relevant affordances in a particular situation. Explaining this selective openness to affordances in dynamical terms should be the main focus for *Radical Embodied Cognitive Neuroscience.* Friston's neurodynamical models provide an interesting perspective on the possible neural mechanisms that underlie the selective openness to relevant affordances. Furthermore, the tendency toward an optimal grip could, within the perspective provided by the FEP, be seen as a consequence of the continuous attunement of the internal dynamics and external dynamics through affordance-responsiveness. We have suggested that the situated anticipation of an affordance generates an action-readiness pattern that makes the affordance stand out as relevant.

Although the FEP alludes to the attunement of the dynamics in the environment and the organism's own dynamics, the body of work in ecological psychology on the rich metastable dynamics in the brain-body-environment system that often is overlooked by the people working on neurodynamics. We have suggested that by bringing together these two approaches with our phenomenology of skilled action, it is possible to develop an integrative research project for understanding affordance responsiveness in both healthy and pathological cases.

From the picture presented in this paper it becomes clear that the common ground for researchers in radical embodied cognitive science, ecological psychologists and dynamical computational neuroscientists is the view that the brain is an intrinsically instable dynamical system embedded in the broader system "brain-body-landscape of affordances". The central way in which an organism relates to its environment is by the organism as a whole tending toward an optimal grip on the field of affordances. It is this phenomenon that should be a central topic in Radical Embodied Cognitive Neuroscience.

In his influential article on action-oriented predictive processing, Clark (2013) is hesitant to embrace the more radical implications of the FEP. In Friston's full free-energy story, one does not need to appeal to desires, goals and rewards in order to explain behavior, but one can replace them with prediction and anticipation; utility functions are replaced by the minimization of prediction error. Clark dubs this resulting picture the "desert landscape" scenario.

However, we think this scenario is rather appealing because unlike almost all work in cognitive neuroscience it does not presuppose the presence of a "goal" or "desire" of which it is completely unclear how it was selected out of the many possible "goals" or "desires." Furthermore, the title "desert landscape" is a misleading depiction of the human ecological niche or landscape of affordances, which is in fact very rich independently of any particular individual. Moreover, we have seen that the skilled individual does not have an explicit goal in mind, but rather is solicited or invited by the field of affordances. We think that what is at the root of skilled activity is not a set of desires or goals, but rather the ongoing modulation of coupled self-organizing dynamical systems that results in the adequate interaction of an organism with its environment.

The best way to characterize these dynamics is in terms of anticipation and attunement. This radical version of the FEP should appeal to enactivists and embodied cognitive scientists, for it does not posit an explanatory role for propositional internal states at the basis of our cognitive system and takes self-organization seriously. Clark's main objection against the desert landscape scenario is that, even if it were true, it will not provide the best way of making sense of the rich organization of our cognitive organization. We agree with Clark here. As we have shown in this paper, an appeal to anticipation and attunement on the subpersonal level in no sense precludes us from highlighting the rich phenomenology of skilled action and the valuable affordances available in the ecological niche. The resulting picture then is not, as it is according to Clark, a barren desert landscape, but rather one of engagement with a flourishing field of affordances.

Chapter 2

The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective[1]

Abstract

In this paper, we argue for a theoretical separation of the free-energy principle from Helmholtzian accounts of the predictive brain. The free-energy principle is a theoretical framework capturing the imperative for biological self-organization in information-theoretic terms. The free-energy principle has typically been connected with a Bayesian theory of predictive coding, and the latter is often taken to support a Helmholtzian theory of perception as unconscious inference. If our interpretation is right, however, a Helmholtzian view of perception is incompatible with Bayesian predictive coding under the free-energy principle. We argue that the free-energy principle and the ecological and enactive approach to mind and life make for a much happier marriage of ideas. We make our argument based on three points. First we argue that the free-energy principle applies to the whole animal–environment system, and not only to the brain. Second, we show that active inference, as understood by the free-energy principle, is incompatible with unconscious inference understood as analagous to scientific hypothesis-testing, the main tenet of a Helmholtzian view of perception. Third, we argue that the notion of inference at work in Bayesian predictive coding under the free-energy principle is too weak to support a Helmholtzian theory of perception. Taken together these points imply that the free-energy principle is best understood in ecological and enactive terms set out in this paper.

---

[1] This paper was previously published as Bruineberg, Kiverstein and Rietveld (2016).

2.1 Introduction: finding a home for the free-energy principle

Anticipatory or predictive dynamics have been at the heart of a large number of influential and very diverse theories of mind, brain and skilled behavior more generally (Helmholtz, 1867; Gregory, 1980; Grush, 2004; Turvey, 1992; Noe, 2004; Stepp & Turvey, 2010; Port & Van Gelder, 1995; Thompson, 2007; Chemero, 2009). These approaches all share a focus on future-oriented activities, but they span a wide range of possible positions in the philosophy of mind. The apparent overlap disguises some fundamental theoretical differences and disagreements. It is within this incompatible plethora of broadly future-oriented frameworks for understanding cognition that one needs to situate the new kid on the philosophical block: the free-energy principle.

The free-energy principle (Friston & Stephan, 2007) is a potentially unifying theory in theoretical neuroscience and theoretical biology, stating that all an organism needs to do in order to maintain its organization as an adaptive living system is to minimize its information-theoretic free-energy in its interactions with the environment. This minimization can be done by predicting or anticipating sensory input or by changing the environment to match what is anticipated. Adequate anticipation requires the organism to be tuned to its ecological niche in such a way that the coupled dynamics of the organism-environment system remain within a relatively small subset of states that maintain the organism's viability in its econiche (Friston, 2011). The promise of the free-energy principle is that it provides a theoretical framework able to unify the biological and the cognitive sciences. The free-energy principle accomplishes this unifying work by showing how the organization and dynamics of living systems prefigures the organization and dynamics of cognitive systems. Living systems that regulate their interactions with the environment so as to maintain their viability, and cognitive systems that are the basis for complex capacities such as social cognition, cognitive control and language use, have fundamental organizational principles in common.[2]

The free-energy principle has often been combined with Bayesian predictive-coding,[3] most notably in the work of its architect, Karl Friston (Friston & Stephan, 2007; Friston & Kiebel, 2009). The main idea of predictive-coding is that the brain generates top-down predictions that are matched bottom-up with sensory information, which results in prediction-errors.[4] According to predictive-coding the-

---

[2] Whether the free-energy principle is able to live up to these theoretical promises is also dependent on its interpretation. We think that an enactive understanding of the free-energy principle, stressing mind-life continuity and self-organization will solve some conceptual problems, which we highlight later.

[3] Also called prediction-error minimization (PEM) by Hohwy (2013, 2014) and predictive processing by Clark (2013, 2015b, 2016a).

[4] For excellent introductions to predictive-coding see (Clark, 2013; Rao & Ballard, 1999; Friston & Stephan, 2007).

orists, the brain is fundamentally in the business of minimizing prediction-errors (Hohwy, 2013; Clark, 2016a). One of the aims of this paper is to better understand the relation between the free-energy principle and Bayesian predictive-coding. For the moment, it is important not to conflate them. The free-energy principle is a unifying framework for self-organizing living systems and is therefore not tied to any particular account of biological or neural functioning. Predictive-coding is a theory of neural functioning. How the latter relates to free-energy minimization is a question to which we will return later.[5]

Friston claims that the principle goes naturally together with older ideas about perception and cognition that date back to the nineteenth century psychologist Helmholtz. These ideas have been influential in more recent years in the work of the perceptual psychologist Gregory (1980) and the cognitive neuroscientist Frith (2013). Helmholtzians take perception to be "unconscious inference", whereby the task of perception is to infer, based on top-down knowledge structures, the current causes of sensation. The process of inference is taken by them to work in ways that are analogous with scientific hypothesis-testing. On the other hand, the free-energy principle is consistent with many other accounts of self-organizing (living) systems such as synergetics (Tschacher & Haken, 2007), global brain dynamics (Freeman, 1987, 2000), metastability (Kelso, 2012; Rabinovich et al., 2008), autopoiesis (Maturana & Varela, 1980) (but see Kirchhoff (2016)) and ecological psychology (Gibson, 1979).[6] Not all of this work is theoretically consistent with a Helmholtzian theory of perception, some of it is even explicitly aimed against such an account. *Our aim in this paper is to show that the free-energy principle and the Helmholtzian account of perception are conceptually independent.* One can explain how the free-energy principle applies to a biological system in terms of the dynamic coupling of the organism with its environment, we suggest. Whether this dynamic coupling is best described using the notion of "inference" we think is an open question. Within the Helmholtzian theory, the notion of inference is standardly understood as a probabilistic relation between prior beliefs, current evidence and posterior beliefs. However, the free-energy principle applies not just to humans but to all living systems, including the simplest of life forms such as bacteria. Friston employs the notion of "inference" even in these minimal cases of simple life forms. This is clearly stretching the meaning of "inference" beyond its normal usage. We question whether the activities of simple life forms such as *E. Coli* can be taken to be inference-involving. In previous work, we have argued for an "ecological-enactive" understanding of the free-energy principle under the umbrella of the Skilled Intentionality Framework (SIF) (Bruineberg & Rietveld,

---

[5] It is important to distinguish between the information-theoretic *quantity* of free-energy and the free-energy *principle*. Predictive-coding is always formally related to the former, but not necessarily to the latter.

[6] Self-organization is a central notion in Friston's work. All these references can also be found in for instance, Friston (2000a, 2013b); Friston, Shiner, et al. (2012) and Friston, Breakspear, and Deco (2012).

2014 (Chapter 1), Figure 1, Rietveld et al., in press; see Ramstead, Veissière, & Kirmayer, 2016; Kirchhoff, 2015; Gallagher & Bower, 2014 for similar arguments). One aim of this paper is to build on our earlier work in part by situating our ideas in relation to those of Hohwy, Friston and Clark, in order to highlight some key differences. We think that Friston's important theoretical work has been mistakenly aligned with Helmholtzian ideas about perception (Hohwy, 2013; Clark, 2013; Friston & Stephan, 2007). The stress on Helmholtz leads to important aspects of the free-energy framework being missed in the philosophical discussion. More specifically, action and perception are not understood in the context of self-organization, and we will argue that this leads to philosophical errors.

Our argument proceeds as follows. In the first part of the paper, we introduce the free-energy principle in more detail. We emphasize the biological motivation for the free-energy principle (as found in Friston and Stephan (2007); Friston (2011)), because it shows that first and foremost the free-energy principle does not apply to brains or epistemic agents, but to embodied living systems as a whole. Second, the biological motivation for free-energy minimization highlights the continuity of mind and life (Thompson, 2007), and hence the overlap with the enactive approach.

In **Section 2.3**, we introduce Bayesian predictive-coding as a theory of brain function and show how it has been taken to form a natural partner with the free-energy principle. We argue against the dominant Helmholtzian interpretation of Bayesian predictive coding in part on the grounds that it provides what we demonstrate to be a problematic conception of the brain as working in much the same way as a hypothesis-testing scientist. In **Section 2.4** we then show how Friston's concept of probabilistic inference, central in much of his work on the free-energy principle, can be given a deflationary interpretation using concepts from dynamical systems theory. We further argue that the free-energy principle and Bayesian predictive-coding constrained by the free-energy principle, contrary to conventional wisdom, provides no support for a Helmholtzian inferential theory of perception.

2.2 The free-energy principle and its enactive foundations

One of the most fascinating questions in the biological sciences is how living systems can produce and maintain their organization in the face of a dynamic environment. A second, equally important and fascinating, question is how biological processes can give rise to minds. The continuity of life and mind thesis, as defended by enactivist philosophers of biology (Jonas, 1966/2001; Di Paolo, 2005; Thompson, 2007), states that both these questions should be tackled at once. Life and mind share the same basic underlying principles. It is exactly this organisational continuity of life and mind that is the starting point of Friston's free-energy principle, suggesting common ground between the free-energy principle and enactive philosophy of biology. Friston agrees that the defining feature of living systems is the way in which "biological systems [...] maintain their states and form in the face of a constantly changing environment" (Friston, 2011, p. 92). Self-maintenance and self-production are the defining features of autopoietic (self-producing) systems, suggesting that there might be an intimate relation between free-energy minimization and autopoiesis as a defining characteristic of life [but see (Kirchhoff, 2016)].

From this conception of life, one can derive a theoretical framework for thinking about perception and action.[7] Free energy is a function of the organism's sensory states and the organism's internal dynamics (called a generative model).[8] Roughly, free-energy is a measure for the ever present dis-attunement between environmental dynamics and internal dynamics (Bruineberg and Rietveld 2014). Free-energy can be minimized on short time scales by making the environment conform to the internal dynamics ("action") or by making the internal dynamics conform to the environmental dynamics ("perception"). This proposal by Friston is, admittedly, a rather unorthodox view of perception. Sensory states here are the proximal stimulation of the organism which can only be changed by acting on the world. Rather than talking about perceptual states we (Bruineberg & Rietveld, 2014 (Chapter 1); Kiverstein & Rietveld, 2015) prefer to speak of states of action-readiness. States of action-readiness are the internal states of the individual that, given its sensory states and abilities, prepare the animal to achieve a grip on a particular situation.[9]

The free-energy principle describes how the very same properties that define life are also essential for cognition, understood as the capacity to regulate interactions with the environment. As Clark has suggested, the free-energy principle may "reveal the very deepest of links between life and mind, confirming and extending the perspective known as "enactivist" cognitive science" (Clark, 2013, p. 24).

Friston formalizes the self-maintenance aspect of autopoiesis in terms of *Shannon information*: of all possible states the organism could find itself in, the organism must find the right subset of states that allow its organization to persist in its energetic exchanges with the environment. The claim is that there is a probability distribution over all possible states the organism can find itself in. At any

---

[7] Taking this starting point is nothing new. A notable precursor in this way of thinking is physicist Erwin Schrödinger. In *What is Life?* (1944) Schrödinger starts from general thermodynamic considerations and develops a theory of metabolism. In an important sense, we take the free-energy principle to be an information-theoretic extension of Schrödinger's work.

[8] The concept of the generative model here should be understood in a particular way as a system of multiple interacting states of action-readiness (see Bruineberg and Rietveld (2014 (Chapter 1))). We expand on this point below.

[9] We expand on this point on p. 79 of the manuscript.

point in time, this distribution is sharply peaked around certain values specifying conditions of the organism that are necessary for its viability and survival or – more exactly – that define the characteristic phenotypic state of the organism. For example, human body temperature has a high probability of being around 37° and a low probability elsewhere. Mathematically, this means that the probability distribution of the variable body temperature has low *Shannon entropy* and that the event 'measuring a body temperature of 37°' has low *surprisal*,[10] while measuring a body temperature of 10° has a very high *surprisal* (for homeothermic organisms like ourselves). Importantly, whether a temperature has low or high surprisal is relative to different species of animal. Birds have a different average body temperature than humans, while ectothermic animals such as lizards have a different distribution altogether.

What holds for internal variables holds just as well for places in the environment. For whales, being in deep sea is an event with low *surprisal*, and being on shore has high *surprisal*, while this is reversed for humans. Hence, the particular embodiment or biological organization of an animal and the environmental conditions of the animal necessary for its viability constrain each other.[11] This is an illustration of the mutualism of animal and environment, taken together they form a complementary pair:

> We commonly talk of the organism and the environment and of the adaptation of one to the other [...] as if there were first an organism and an environment and then some adjustment of one to the other; but when we come to an analysis of the factors involved, it is quite necessary to start from the unity of function and see that the distinction of organism and environment arises because of adaptation in that process, not vice versa. (Dewey, 1976, p. 275, as quoted by Costall, 2004; see also Maturana & Varela, 1980)

So far, all we have said is that the embodiment of the animal implies a range of environmental conditions in which it is able to prosper and that surprisal is a measure that captures this relationship. As such, the free-energy principle is just an information-theoretic formalization of the observation that living systems are homeostatic (Bernard, 1865/1927) or homeodynamic (Yates, 2008) systems. What is required is an account of *how* living systems are able to find the right

---

[10] Shannon entropy is the expected or average surprisal and surprisal is related to an event with a particular outcome.

[11] In forms of life with sociocultural practices this environment is highly complex and best understood as a sociomaterial environment (Mol, 2002). The FEP is not limited to understanding adaptation to an environment offering resources for meeting basic biological needs such as food, but can also be applied to adaptation to the practices found in the human social environment. The landscape of affordances in the human form of life is very rich (Rietveld & Kiverstein, 2014; Bruineberg & Rietveld, 2014 (Chapter 1), Figure 1).

subset of states so as to maintain themselves within viable bounds.[12] The ideal temperature of a human is determined by its embodiment: at 37° the enzymes regulating our metabolism perform optimally, while the metabolic cost of maintaining body temperature is still manageable within certain environmental conditions. The challenge that the homeostatic system faces is that of regulating the animal's interactions with the environment so as to keep the animal within this viable regime, without "knowing" about the animal's own viability conditions. Surprisal is relative to this unknown probability distribution, thus an organism has no means of evaluating surprisal directly. Since surprisal is a more general term in information-theory, we will call this special case of surprisal, only related to the embodiment of the agent, *embodied suprisal* or *surprisal*$_E$. We will call the unknown distribution, which formally describes the conditions of viability of the animal, the *bodily distribution.* How then does the organism succeed in finding the right subset of states?

This is where the free-energy principle may hold a solution. The main insight of the free-energy principle is that information-theoretic free-energy is always greater than (or equal to) surprisal$_{\mathrm{E}}$[13] and is thus an upper bound on surprisal$_{\mathrm{E}}$. By minimizing free-energy, the organism will implicitly minimize surprisal$_{\mathrm{E}}$ as well, and hence remain within viable bounds. Free-energy is accessible to the organism because it is a function of two quantities (i.e., probability distributions) that the organism embodies: a generative model and a variational density that is entailed by a system's internal state. Mathematically, the generative model is thought of as the probability of the co-occurrence of a sensory state and a state of the environment. The variational density is a proxy for the real bodily (environmental) probability distribution (e.g., the distribution of temperature). Typically, one assumes that this distribution is Gaussian; making it fully specifiable by only its mean and standard deviation. Changing the variational density (i.e., changing its mean or standard deviation) will give different values for the free-energy. What is needed then for free-energy *minimization* is for the organism to be able to change the variational density or to change the environment in a way that reduces free-energy. Friston appeals to the derivative of free-energy with respect to internal states (that encode the variational density) and to sensory states.[14] This implies nothing more, we take it, than that the organism regulates

---

[12] In this section, we focus on biological examples to highlight the link between homeostasis, self-organization and the free-energy principle. Ultimately, the free-energy principle should help us to understand all improvements an individual makes in a situation. We will come back to this in the last section.

[13] Formally, free-energy is defined as surprisal$_{\mathrm{E}}$ plus the KL-divergence between the recognition density and the embodied distribution. Since, KL-divergence is strictly positive, free-energy is an upper bound on surprisal$_{\mathrm{E}}$.

[14] We explain what Friston means by an internal state later in this paper when we introduce the notion of a Markov blanket.

its interactions with the environment and its own internal conditions based on what it takes to be the norm of maintaining its own viability or flourishing more generally.

We argued so far that surprisal$_{\mathrm{E}}$ (for instance related to temperature) cannot be evaluated directly because the 'true'[15] embodied probability distribution is unknown. Instead, based on the generative model the system itself generates a variational density about its own optimal temperature that can be used to evaluate "how far off" the current temperature is. We will use the general information-theoretic notion of surprisal to describe this discrepancy. Free energy is always greater than the average surprisal$_{\mathrm{E}}$. This can be seen intuitively: if the internal mechanism regulates temperature to be 38° , while our embodiment requires 37°, there will be a big discrepancy ('surprisal$_{\mathrm{E}}$') over time. The closer the variational density approximates the bodily distribution, the lower the surprisal$_{\mathrm{E}}$ will be. Hence minimizing the discrepancy between the 'predicted' and the 'actual' temperature requires that the following condition be met: the internally generated (variational) distribution and the actual distribution must be mutually compatible with respect to the sensory constraints (in the sense of being sufficiently similar enough in their dynamics).[16]

The organism changes the variational density by changing its internal dynamics with respect to the constraints of the environment. Friston calls this "perceptual inference". The latter is not sufficient for minimizing surprisal$_{\mathrm{E}}$ (and therefore survival, if the logic of the free-energy principle holds). In order to minimize surprisal$_{\mathrm{E}}$ an organism needs to change its sensory states, which it does through action. Perception, understood in Friston's terms as changing the internal dynamics of the organism, will only lower the upper-bound on surprisal$_{\mathrm{E}}$ (i.e., free-energy) and not change surprisal$_{\mathrm{E}}$ itself. Crucially, to change *surprisal$_E$* itself, the organism needs to change its sensory states *by acting* on its environment. Suppose a human being finds itself with a temperature of 39.8°. If it would be accommodating enough to change its internal dynamics so that this was its expected temperature, it would not survive for long. What matters for the organism is avoiding finding itself in such a state by changing its sensory states through acting in the world (e.g., by going to the doctor).

Action is therefore necessary for minimizing surprisal$_{\mathrm{E}}$; this cannot be done on the basis of perception alone since perceptual inference cannot reduce surprisal$_{\mathrm{E}}$. Mathematically, this can be seen by the fact that one can write the free-energy

---

[15] The 'true' distribution is a technical term in statistics. The relevant distinction is that between a 'true' and an 'empirical' distribution. In this context, the 'empirical' distribution is the variational density that estimates the 'true' embodied distribution.

[16] Of course this assumes that the generative model is sufficiently in accordance with the bodily distribution. This requires an evolutionary and developmental explanation that takes into account the variability of the animal's niche or the human form of life. In itself, the notion of free-energy minimization is applicable also on phylogenetic and ontogenetic timescales (Friston, 2011, Figure 10).

term as a Kullbeck–Leibler[17] divergence between the variational density and the posterior plus the term for surprisal$_{\mathrm{E}}$. Minimizing this term by only changing the internal states (perception) minimizes the KL-divergence, and not the surprisal itself (see **Figure 2.1**, see also for instance Friston, 2011, Figure 1). We will in the following refer to KL-divergence as "divergence".

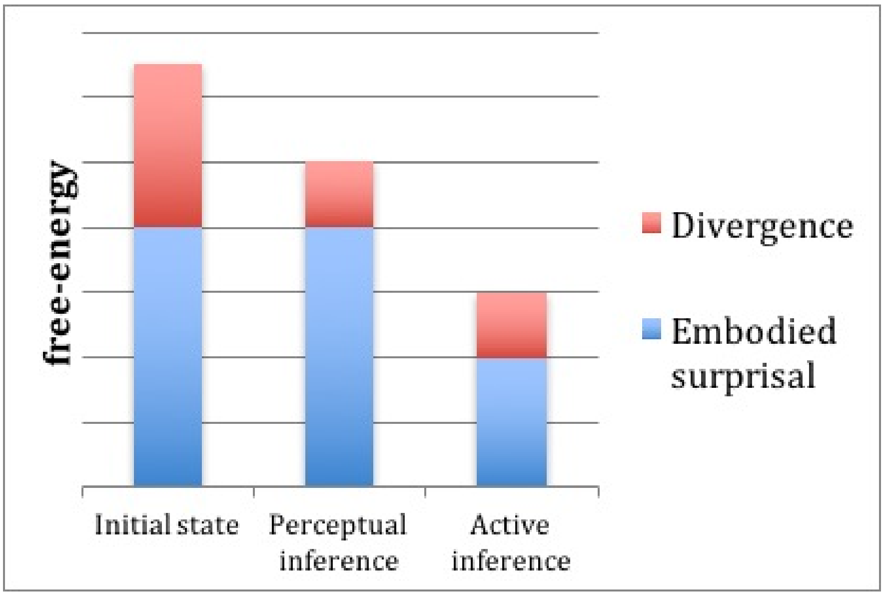


Figure 2.1: Schematic depiction of how free-energy is composed of divergence and embodied surprisal (surprisal$_{\mathrm{E}}$) and how these components change with perceptual and active inference respectively (see text). Perceptual inference can only reduce divergence, but not embodied surprisal. Only active inference can reduce embodied surprisal.

So within the free-energy framework, it is *action* that does the work of actually minimizing surprisal. Actions change an organism's relation to the environment, thereby changing the sensory states of the organism, a process that Friston calls active inference (2012). Free energy, as we understand it (Bruineberg and Rietveld 2014), is a measure of the dis-attunement of the internal dynamics and the environmental dynamics: it is low when the sensory states are anticipated, by the animal, and high when they are not. The free-energy principle says that minimizing free-energy is a necessary[18] condition for living systems to maintain

---

[17] The Kullbeck-Leibler divergence is the "extra" surprisal obtained for approximating the *bodily distribution* with a generative model. Intuitively, when throwing a fair dice, there is an expected surprisal per throw. When falsely expecting the dice to be skewed, there is on average an extra amount of surprisal gained. The difference between the two is the KL-divergence.

[18] Free-energy minimization is sufficient for self-maintenance in the sense that it guarantees the system to frequent an invariant set of states. However, this is not sufficient for a demarcation of living from non-living systems (since a simple homeostat meets these criteria). We believe additional (enaction based) conditions are needed (see for instance Froese & Stewart, 2010). We thank reviewer 2 for pressing us on this point.

their organization in their econiche. What is crucial for the organism, is that it anticipates the kind of interactions with the environment that contribute to its flourishing (such as for example eating food, staying warm, avoiding a passing car or learning to read). The function of the generative model is therefore not to provide the agent with a representation of the dynamical structure of the environment *per se*. It mediates the organism's interactions with its environment in such a way that a *robust* brain-body-environment system is maintained. The generative model is not a model of the environment as it happens to be. It embodies in its organization and structure longer-term regularities between action, environment and the state of the organism. How can we understand the relationship between the niche and the generative model?

Friston claims that "an agent does not *have* a model of its world, [but rather] it *is* a model" (Friston, 2013a). One thing this means is that the free-energy principle does not just apply to the brain, but rather to the organization of the biological agent as a whole in its econiche. Friston himself states, somewhat provocatively, that: "each [...] agent embodies an optimal model of its econiche" (Friston, 2011, p. 89). In previous work (Bruineberg and Rietveld, 2014) we have provided a minimal yet still practically useful interpretation of '*being* a model of its econiche'. Minimally this implies: (1) that the agent embodies in its structure and organization skills that complement the econiche and (2) that the coupled dynamics of organism and niche[19] together lead towards some dynamical equilibrium or grip on the situation. The free-energy principle provides an explanatory framework for understanding the coupling, or attunement, between internal and external dynamics. Following Friston, we are led to the conclusion that, because of the organism's skills and affordance-related states of action-readiness, the organism itself *is* its own best model of the causal structure of its niche, a paradoxical sounding twist on the familiar embodied theme that "the world is its own best model" (Brooks, 1991).

The main motivation for the free-energy principle is the biological requirement for self-organization and self-maintenance. The causal processes that hold internal variables and environmental variables within bounds are coupled. For instance, a decrease in blood sugar level of a mammal might alter the anticipated sensory input such that it anticipates finding a food source, hence making the behaviour of the animal conform to its metabolic needs. Auletta (2013) provides a good example of how a coupled informational (sensorimotor) and metabolic system can

---

[19] Econiche, here and elsewhere in the paper, is understood as the landscape of available affordances to a certain animal species or form of life, see (Bruineberg & Rietveld, 2014 (Chapter 1), Figure 1).

provide a model for bacterial chemotaxis.[20] The system that is minimizing free-energy is nothing less than the whole self-organising and self-regulating affordance responsive system.

The free-energy principle then reveals the deep continuity between mind and life that Clark (2013) has hinted at as being reminiscent of enactive approaches towards life and mind (Varela et al., 1991; Thompson, 2007; Stewart, Gapenne, & Di Paolo, 2010; Colombetti, 2014; Di Paolo, Rohde, & De Jaegher, 2010).[21] From our enactive point of view, it is the myriad of self-maintaining and self-producing processes that make up the animal, that give the animal a lived perspective on its environment. Metabolic and thermal disequilibria structure the way in which the world is perceived in the sense that it makes the individual organism selectively open and responsive to *relevant* affordances, to the relevant possibilities for action the environment offers. An apple might invite eating when I am hungry, and a blanket might solicit wrapping around me when I am cold. What counts as minimizing surprisal is thus relative to the current state and situation of the animal and its lived perspective on the many affordances offered by its environment. In acting in order to minimize surprisal, the individual is responsive to only the disequilibria with the environment that are relevant given its current state and situation.

In many kinds of animals, these disequilibria might be experienced as an affective tension or something-to-be-improved. The animal is drawn to act by the relevant affordances in its situation, so as to reduce this tension, which results in its improving its grip on its environment. We call this the tendency towards an optimal grip on the situation, which we take to be a basic concern of every living animal (Rietveld, 2008a; Bruineberg & Rietveld, 2014 (Chapter 1); Kiverstein & Rietveld, 2015). This basic concern to improve grip is a precondition for and structures the animal's lived perspective on the basis of which certain affordances offered by the environment stand out as significant or relevant. Minimizing surprisal then can be understood as an animal reducing its disattunement with the environment. In doing this, the animal improves its grip on the environment.

There are two points we therefore wish to stress that will be important for our later argument. First, surprisal is relative to an animal with an affective perspective on its ecological niche. The individual cares about something, namely about improving or at least maintaining grip on its changing situation. Friston emphasizes self-maintenance as a distinguishing feature of living systems from other self-organizing systems, where self-maintenance is understood as the ability to change their relationship with the environment and maintain thermodynamic homeostasis (Friston & Stephan, 2007, p. 5). We have characterized this abil-

---

[20] In fact, Auletta (2013) mentions the genetic system as a third subsystem that constrains both the metabolic and the sensorimotor system over long time scales. Since we only deal with perception and action in this paper, and not with development and evolution, the genetic system should not concern us here.

[21] See quote on p. 69.

ity as an instance of what in philosophical phenomenology is described as being moved so as to tend towards an optimal grip. Thus on our view self-maintenance implies an affective perspective on its environment. Certain aspects of the animal's econiche stand out as significant for the organism on the basis of its basic concern to improve its grip on the environment, and to reduce its disequilibria with the environment.

Second, we have argued that free-energy minimization is best understood in terms of the disattunement of the internal and external dynamics of the animal and its environment. It is the whole organism in regulating its interactions with the environmental landscape of affordances that is minimizing surprisal. We will see in the next section, how Jakob Hohwy argues that minimizing surprisal is the evolutionary function of the brain. Hohwy takes the minimization of surprisal to be a function of the brain in the same sense as the function of the heart is to pump blood. We've argued by contrast that minimization of surprisal is done by the whole organism being drawn to act on relevant affordances in ways that result in the reduction of dis-attunement with the environment.

2.3 The Helmholtzian brain as a crooked scientist

Jakob Hohwy has drawn on the free-energy principle to develop a theory of the predictive brain that is very different from the ecological and enactive one we have just proposed. He interprets the free-energy principle using older ideas from Helmholtz (1867) and Gregory (1980), following up on suggestions that can also be found in Friston's writings (see for instance Friston & Kiebel, 2009). Hohwy rightly points out that the Helmholtzian theory of perception is strongly internalist in a way that challenges enactive and ecological approaches towards the mind. In *The Predictive Mind* (2013), Hohwy has given one of the most comprehensive and careful expositions of the prediction-error minimization framework (PEM) to date. In a more recent follow-up paper, he has the following to say about what his interpretation of PEM implies for a philosophical account of the mind-world relationship:

> PEM should make us resist conceptions of this relation on which the mind is in some fundamental way open or porous to the world, or on which it is in some strong sense embodied, extended or enactive. Instead, PEM reveals the mind to be inferentially secluded from the world, it seems to be more neurocentrically skull-bound than embodied or extended, and action itself is more an inferential process on sensory input than an enactive coupling with the environment. (Hohwy, 2014, p. 1)

Generally in formulations of PEM, brain processes are taken to work according to a unifying principle related to but not identical with the free-energy principle. PEM claims this unifying principle is the minimization of prediction-error between the actual inflow of sensory input occurring over multiple temporal and spatial scales, and the flow of sensory input the brain predicts based on its top-down knowledge. Sensory processing in the brain is not a bottom-up process of feature detection but instead consists of the top-down prediction of sensory signals based on hierarchically organised models of environmental causal regularities. What gets conveyed through bottom-up processing are weighted prediction errors that arise when there is a discrepancy, or "prediction-error" between incoming inputs, and what the brain expects. Sensory processing thus becomes about predicting sensory input. By minimizing prediction-errors over time (given a suitable generative model) the system "will come to infer the hidden state of the environment".

Hohwy equates PEM with the view of perception as knowledge-driven perceptual inference as advocated by Helmholtz and Gregory. The logic is the following: the brain only has access to its own predictions and to prediction-errors and based on these it has to infer the hypothesis that best explains away current prediction error. Perception then essentially is hypothesis-testing in which the brain seeks to explain away prediction-errors by finding the hypothesis that best explains the available sensory evidence. The 'unconscious inferences' that lie between a sensation and a perception are much like the inferences that scientists draw (Gregory, 1980; Hatfield, 2009). It is the inferred hypothesis about the causes of sensory input that provides our epistemic access to the external world. The brain only has access to the ways its own patterns of neural activity and spike trains flow and alter. The true states of the external environment are hidden from the brain beyond the veil of sensory information. Any epistemic access to the world is therefore, according to Helmholtzians like Hohwy, indirect based on the hypothesis that best explains away current prediction error.

This Helmholtzian line of reasoning assumes however a distinction between sensation and perception that is problematic for ecological and enactive cognitive scientists. The general disagreement between inferential and ecological approaches to perception concerns the richness or sparsity of our sensory contact with the environment. For the Helmholtzian, sensations are impoverished and need to be enriched by top-down knowledge before the perceiver can know what is out there in the world. This view of perception as starting from impoverished and ambiguous sensations does not fit with the richness of the ecological context. For the ecological theorist, the perceptual system is attuned to directly (i.e., non-inferentially) pick up on environmental regularities in the organism's niche. Within that framework, perception is understood in terms of lower-level phenomena like resonance and attunement, mediated by the organism's skills.[22]

As we have argued for above, the animal species and the ecological niche are co-specifying. The niche of the animal is best understood as a landscape of affordances that reflects the perceptual capacities and other abilities of a kind of animal belonging to a particular form of life (Rietveld & Kiverstein, 2014). Affordances are more precisely defined as relations between aspects of the material environment and abilities available in a 'form of life' (ibid., p. 337).[23]

The lived perspective of an individual organism reflects its needs, concerns and abilities and determines which affordances are relevant in a particular situation. Hohwy is absolutely correct to pick up on a tension between Helmholtz and conceptions of the mind like our own that stress openness to the world. However, unlike Hohwy we take the free-energy principle to challenge those that would follow Helmholtz.

Hohwy conflates PEM with the free-energy principle. He writes: "Since the sum of prediction error over time is also known as free-energy, PEM is also known as the free-energy principle (Friston & Stephan, 2007)"(Hohwy, 2014, p. 2). We believe this is confusing, if not simply wrong. There is a difference between the free-energy *bound* and the *free-energy principle.* The free-energy bound is the computational trick of reducing an (intractable) integration problem to an optimization problem by introducing a variational density.[24] The free-energy *principle*, on the other hand, is tied to the minimization of surprisal and therefore (by the logic of the free-energy principle) to self-organization and what we call the tendency towards an optimal grip on affordances. As we have shown in **Section 2.2**, one important consequence of the free-energy principle is that it is *only action that minimizes* $surprisal_E$. Perception, understood as changing the internal dynamics of the organism, will only lower the upper-bound on $\mathrm{surprisal}_\mathrm{E}$ and not change surprisal itself.

For Hohwy, perceptual inference is changing the hypothesis so that it fits the data, whereas active inference is changing the data so as to fit the hypothesis. Both are different strategies for minimizing prediction-error. Hohwy has a par-

---

[22] As Orlandi (2016) points out, there is an interesting historic continuity between Gestalt psychology, ecological psychology, connectionism and predictive-coding, whereas there is a discontinuity between predictive-coding accounts and a Helmholtzian theory of perception. We think that this continuity between ecological psychology and predictive-coding is important and becomes undeniable in the context of the free-energy principle.

[23] Note that the human form of life includes socio-cultural practices and many of the affordances of the human (sociomaterial) environment and the corresponding abilities have a place within these practices.

[24] This method is used extensively in machine learning under the heading of variational inference or variational Bayes (for an introduction, see Murphy (2012), Chapters 21 and 22). This technique is mainly used in forms of inference where one wants to investigate which causal structure most likely generated the data.

ticular way of adding action to perceptual inference.[25] In perception the brain tests its current hypothesis of the world by predicting the sensory input or data it would expect on the assumption that its current hypothesis is true. It then compares the predicted data with the actual data and goes on to adapt its hypothesis to improve the fit with the actual data. The brain laboriously changes and fine-tunes its hypothesis as experimental results come in. These iterations of hypothesis-testing are what Hohwy has in mind when he talks about *perceptual inference*. He introduces active inference by appealing further to the analogy between perception and scientific hypothesis testing (Hohwy, 2013, p. 77). Scientific hypothesis testing is not just passively recording results of an experiment, but carefully setting up experiments and actively intervening in the chains of causes and effects in order to disentangle the causal web. In action the brain controls movement, sampling the environment with the aim of matching actual sensory input with predicted sensory input. According to this framework, the role of action is to support perceptual inference and makes predictions more reliable (Hohwy, 2013, p. 79).

However, we have seen that form the perspective of the free-energy principle, *perceptual inference without active inference does not make sense*, since only action minimizes surprisal. Let us return to the example of temperature we used in **Section 2.2**. If you find yourself too hot, one way to reduce your temperature is by taking a cold shower. *Action* here is doing the work of reducing surprisal$_{\mathrm{E}}$, because of its effect on the animal's relation to the environment, taking it closer to an optimal bodily condition. Perceptual inference also works in the service of actively improving the organism's grip on the environment. We've followed Friston in characterising this form of inference in terms of change in the animal's internal (endogenous) dynamics. Perceptual inference can thus be thought of as patterns of action-readiness. In the case of the shower, the sensation of being too hot shapes your perspective on the environment in such a way that the cold shower now stands out as inviting or attractive (whereas normally it probably would not). Perceptual and active inference are not two distinct strategies for minimizing prediction-error (each with their own so-called 'directions of fit'), as Hohwy suggests. Perceptual and active inference are entangled parts of a single process of readying the organism to act in such a way as to improve its grip on the environment. Perception, understood as changing the organism's internal dynamics reflects the organism's pattern of action-readiness and thus its selective openness to affordances. The organism is ready to act on relevant affordances so as to improve its grip on the environment thereby reducing surprisal. There is no separation between perception and action because perception on our account just is the organism's preparing itself to act in ways that reduce surprisal, thereby

---

[25] As originally conceived, active inference (Friston, Mattout, & Kilner, 2011) subsumes both perception and action. It purports to generalize the Helmholtzian perspective to include action, rendering them inseparable. We will return to this later.

improving grip on its environment both by getting ready for what is to come and by engaging with affordances offered by it in action. The environmental affordances scaffold the individual's actions. Action does not have a world-to-mind direction of fit, as Hohwy argues, because in acting the agent has no intention or goal in mind. Instead the agent is drawn to act based on its disattunement with the environment. It is the affordances relevant for reducing this disattunement that drive the actions of the agent.

Both Hohwy and Friston acknowledge the central role for action in the minimization of surprisal (Hohwy, 2013, pp. 84-88), but we think they fail to realize the damage it does to their Helmholtzian commitments. As mentioned earlier in this paper, on a Helmholtzian account of the mind, the aim of perceptual inference can be said to be to infer the most likely cause of sensory input, the more objective the better. However, we would argue that a perfect hypothesis that precisely represents the state of the environment is worthless if it does not specify what action minimizes surprisal, or improves grip. Hohwy recognizes this point when he writes as follows: "perceptual inference can make you perceive that you are hurtling towards the bottom of the sea with something heavy around your feet but cannot do anything to change that disturbing sensory input which is fast taking you outside your expected states." (Hohwy, 2013, p. 85) It is only by action that you can do something about this disturbing sensory input by, for instance, untying the heavy object from your feet. However, active inference of the type involved in such a life or death situation, would be ecologically useless if it did not predict adaptive actions that would improve my situation in this life threatening environment. The possibility of cutting the rope with my Swiss army knife in my pocket should become the relevant affordance that drives my actions now.

The upshot of this is that my brain is not, and should not behave like an exemplary scientist. If my brain really is a scientist, then it is heavily invested in ensuring the truth of a particular theory, which is the theory that "I am alive". This is a fundamental prior belief that drives all action; namely, I exist and I will gather all the evidence at hand to prove it. It will only make predictions whose confirmation is line with this hypothesis. It does not give competing hypotheses a fair chance and is extremely biased in the way it interprets the data. It decides on the outcome of an experiment beforehand (my staying alive) and manipulates the experiment until the desired result is reached. If my brain is a scientist, it is a crooked and fraudulent scientist–but the only sort of scientist that can survive an inconstant and capricious world.[26] The hypotheses that the brain in reality is biased in favor of, are hypotheses that predict the animal will tend towards grip on environmental affordances.

---

[26] Thanks to Karl Friston for suggesting this formulation.

The interesting question in active inference is therefore not, as Hohwy claims, how the brain can use available sensory input to accurately reconstruct the hidden state of affairs in the world (Hohwy, 2014, p. 1). The interesting question is rather how the space of possible 'hypotheses' is always already constrained in such a way as to make the animal improve its grip on the environment.

Hohwy might seem to recognize this point when he discusses the limitations of actions understood as active sampling of sensory input. However, he attempts to offer an account of PEM that abstracts away from questions about how the space of hypothesis is already constrained: "I will not engage these types of questions directly here. I am primarily interested in what the account says about our understanding of the world and our place in it as perceivers and agents" (Hohwy, 2013, p. 88). We, however, think that this is where the free-energy principle starts to get interesting: one cannot understand our place as agents in the world without taking into account self-organization. The free-energy principle with its focus on self-organizing dynamical systems is fundamental to answering these questions about the mind-world relation. The Helmholtzian metaphor of the brain as hypothesis-testing scientist is ill-equipped to capture active inference under the free-energy principle for the simple reason that it triggers the wrong questions. The central question for the organism is not what the state of affairs of the environment is (although this might be relevant for achieving certain forms of grip), but rather which actions will minimize surprisal, where surprisal is taken relative to an encultured, skilled and embodied agent. The organism is always acting on the basis of a basic concern to improve grip, and this biases and informs all of its actions. This basic concern is taken for granted or presupposed by the Helmholtzian.

This presupposition shows up for instance in the way in which active inference is informed by and depends upon the goals and intentions of the agent. For example, suppose I am falling down the stairs. I could perfectly predict myself tripping down the stairs and breaking my neck. Navigating the stairs is seen by the Helmholtzian as the goal of the agent, and relative to that goal active inference works so as to bring about perceptions that take you closer to accomplishing that goal. However, it remains a mystery why the agent has selected the goals that it has in this situation. Implicit in this appeal to goals and intentions is the basic concern of the organism towards improving its grip on its environment. The point *per se* is not about prediction-error minimization, because my brain can be correctly minimizing prediction-error as I fall down the stairs. Although my sensations might tell me it is very likely that I am tripping down the stairs, my brain *needs* to treat this as a highly unlikely event and do as much as possible to return to a more likely situation (i.e., balancing on two feet). These considerations are echoed in Friston's (2011) account of embodied inference:

> I model myself as embodied in my environment and harvest sensory evidence for that model. If I am what I model, then confirmatory evidence will be available. If I am not, then I will experience things that are incompatible with my (hypothetical) existence. And, after a short period, will cease to exist in my present form. (p. 117)

The Helmholtzian is thus already taking for granted without explaining something which is central to our account, the lived perspective and the concerns of the organism. Rather than simply presupposing goals and intentions, we suggest that the basic concern of the organism is tending towards improving its grip on available affordances in its environment.

Unlike Hohwy (2014) and Clark (2013)[27] we follow Friston in taking the free-energy principle to be the biological principle of living systems as a whole, and not only of the brains of living systems. Goals and intentions ought to not be something external to the free-energy principle, but rather should be explained by the principle itself. The Helmholtzian account does not in itself provide any explanation of the goals of the organism, and why the organism selects the goals it does. It rather specifies prediction-error minimization in relation to those presupposed goals without offering any explanation of the process by which the goals themselves are formed. By contrast our account understands free-energy minimization in terms of the tendency towards an optimal grip on available affordances. It thereby shows how selective openness to relevant affordances follows from acting according to the free-energy principle. There is therefore no need to presuppose goals and intentions whose origins remain quite mysterious, or for an additional account of goal and intention formation.

Our concept of the tendency towards an optimal grip additionally puts the Bayesian notions of precision and uncertainty in a different perspective.[28] The traditional function of uncertainty in Bayesian accounts of perception is to modulate the impact of sensory perturbations (‘prediction-errors’) on the internal dynamics (‘hypothesis’). If the agent has high confidence in its sensory input and low confidence in its current hypothesis, then the sensory prediction-error will greatly shift the hypothesis. In the reversed case (low sensory confidence and high internal confidence), the probability density will be unaltered.

Confidence however cannot be a function of sensory evidence alone taken in isolation from contexts of skilled action and engagement with the world. The central question for the agent and thus for active inference is not to settle on which hypothesis is true, but on *what needs to be done.* Even as all sensory

---

[27] Since his 2013 paper, Clark seems to have changed considerably his position regarding the relation between embodied cognition, Helmholtzian perception and predictive-coding. In recent work he (Clark, 2015b), is no longer committed to an inferential Helmholtzian view of perception, but rather talks about grip. However, also in this work, it remains unclear where goals and intentions come from.

[28] We thank an anonymous reviewer for urging us to make this explicit.

evidence points to me standing under a shower that is too hot, I *need* to treat this as an unlikely event and arrange my *actions* accordingly. Moreover, most of our actions are extremely context-sensitive: the ringing phone solicits answering when I am alone, for example, but not when I am having a conversation. Situations typically offer multiple possibilities for action, and the degree to which each of these possibilities stands out as salient or relevant is due to precision-modulation.[29] Crucially, such precision-modulation is structured by an agent's skills and concern–their acquaintance with a normative socio-cultural practice (e.g., when it's acceptable to answer your phone); their habits (getting a cup of coffee before starting work); their bodily needs (eating an apple when hungry). While engaged in a conversation a buzzing phone leaves us cold (does not alter internal dynamics, has low precision), while in another context it solicits answering (high precision, impacts internal dynamics). Precision-modulation based on these kinds of factors, shapes the salience and relevance of the field of affordances with which agents engage. We cannot understand our place as agents in the world without taking into account the wider contexts and situations in which skilled action takes place.

2.4 Inference behind a Markov blanket

Both Hohwy (2014) and Clark (2013) take the brain to occupy an epistemic position that implies a Helmholtzian theory of perception. They borrow from Eliasmith (2005) the notion of a perspective, to be able to specify for each neuron or brain mechanism what information it has access to in a strict sense:

> A 'perspective', as I shall use the term, is a relation between an information processor and a transmitter of information. Perspective is determined by *what* information is available to an information processor from a transmitter (Eliasmith, 2005, p. 99, italics in original)

The notion of information implied here is just information in terms of (meaningless) energy transfer and should not be confused with intentional or semantic content. As Eliasmith puts it, the 'perspective' is a simple relation between a transmitter and a receiver of information. Assigning such perspectives is a useful tool for tracking information flows and covariance relations in complex systems like the brain. Eliasmith goes on to make the trivial sounding claim that "[A]n animal (and each of its information processing sub-components) can only access information available through sensory receptors" (Eliasmith, 2005, p. 100, italics in original). It is now a short step to the conclusion that since all the brain has direct access to are the changes that are taking place in its sensory registers,

---

[29] For our positive account of how this relevance sensitivity might work see the section 'Neurodynamics of selective openness' in Bruineberg and Rietveld (2014 (Chapter 1)).

the only route to knowledge of what it is in the environment that is causing its changing sensory input must be through inference of some kind. Clark (2013) captures this type of reasoning in the following passage. We quote him in full since it is precisely this line of reasoning that we wish to challenge in this final part of our paper. Clark writes:

> For, the task of the brain, when viewed from a certain distance, can seem impossible: it must discover information about the likely causes of impinging signals without any form of direct access to their source. Thus, consider a black box taking inputs from a complex external world. The box has input and output channels along which signals flow. But all that it "knows", in any direct sense, are the ways its own states (e.g., spike trains) flow and alter. In that (restricted) sense, all the system has direct access to is its own states. The world itself is thus off-limits (though the box can, importantly, issue motor commands and await developments). The brain is one such black box. How, simply on the basis of patterns of changes in its own internal states, is it to alter and adapt its responses so as to tune itself to act as a useful node (one that merits its relatively huge metabolic expense) for the origination of adaptive responses? Notice how different this conception is to ones in which the problem is posed as one of establishing a mapping relation between environmental and inner states. The task is not to find such a mapping but to infer the nature of the signal source (the world) from just the varying input signal itself. (Clark, 2013, p. 3)

Clark seems to imply in this passage that there is a strong epistemic boundary separating the brain from the outside world. The brain has to infer the hidden state of the world that lies beyond the veil of sensory information. Any epistemic access to the world has to be inferential, because the brain has no direct access to the causes of its sensory information. Based on the flow and alteration of its spike trains, the brain needs to internally reconstruct the causal structure of the world outside. This would seem to challenge the openness of the organism to the environment, suggesting instead that the organism is secluded and cut-off from the environment.[30]

---

[30] In his recent work, Clark argues that there is no inconsistency between PEM and embodied and enactive approaches towards the mind (Clark, 2015b, 2015a). He sees no incompatibility between a theory of the brain as a prediction engine and the openness of the organism to its environment. Clark has argued against a reconstructive account of perception (such as that of Helmholtz) in which the task facing perceptual systems is to form an objective representation of the external world and its causal structure. He argues instead that sensing is inseparable from moving and that the organism acts so as to generate the sensory input it expects, coordinating its behavior with relevant aspects of the distal environment. Clark draws on Michael Anderson's

We will argue that the boundary between the organism and the environment is real, but it is not the kind of boundary that hides causal structure in the environment from the organism.[31] We make this argument from simulations of very minimal cases of life and cognition as discussed by Friston (2013b). In this paper, Friston shows two things. First, that coupled ergodic dynamical systems with a Markov blanket are a sufficient condition for free-energy minimization and second, that such systems will spontaneously occur in a simulation of a "primordial soup" with relatively arbitrary assumptions.[32] The Markov assumption, as it is known from statistics, is a conditional independence of a node with its predecessors, given its immediate predecessors (or "parents"). For instance, in a domino chain reaction, the behavior of a given tile is only dependent on the behavior of its predecessor in the chain. Although the behavior of the predecessor is dependent on the rest of the causal chain, given the behavior of its predecessor, the domino of interest and the rest of the causal chain are statistically independent (**Figure 2.2**).

The Markov blanket is an extension of the Markov assumption in the sense that it makes a node or a subgraph conditionally independent from the rest of the network given its neighboring nodes or 'Markov blanket' (Pearl, 1988). Markov blankets are mainly used in the context of Bayesian graphical models, but as Friston (2013b) shows, they are all pervasive in natural systems dominated by short-range interactions. Due to the Markov blanket, Friston argues, a random

---

(2014) idea that every perception is accompanied and preceded by some action. Perception can be thought of as opening a sensory channel to the world, and action as controlling the inputs the perceiving animal receives via this sensory channel.

[31] The recent literature (Clark, 2013) on Helmholtzian interpretations of predictive processing contains a number of commentaries from enactivists that overlap with the arguments of this paper in suggestive ways. Froese and Ikegami (2013) criticize Clark (2013) for failing to exclude agents that pursue the "dark room solution", seeking out a maximally predictable but ecologically unsatisfying environment. Our "crooked scientist argument" (see **Section 2.3**) can be read as a further extension of this enactive intuition: prediction- error minimization is not the *ultimate aim* for the organism, but rather a *means* by which a better grip is achieved. We suggest that it is only if the generative model embodies the econiche of the organism that minimizing prediction-errors will lead to ecologically relevant outcomes. In another commentary Roesch et al. (2012) criticize Clark (2013) for not doing justice to the affective perspective of the organism. The account of the free-energy principle presented in this paper explicitly takes the affective perspective of the organism as its starting point (see p. 75 of this dissertation). The solicitation or attractiveness of relevant affordances, which is central in our account, has an affective character.

[32] The "primordial soup" simulation in Friston (2013b) consists of the interaction between a number of particles with Newtonian and electrochemical dynamics. When these particles interact with each other over time, the system tends towards a stable configuration in which the system is describable in terms of "internal", "external", "perception" and "action" states, based on their causal dependencies. Although not directly causally coupled, the electrochemical dynamics of the internal and external states of the system become synchronized. On Friston's interpretation, this allows an understanding of the internal dynamics as inferring the hidden dynamics of the environment based on the sensory and action states.

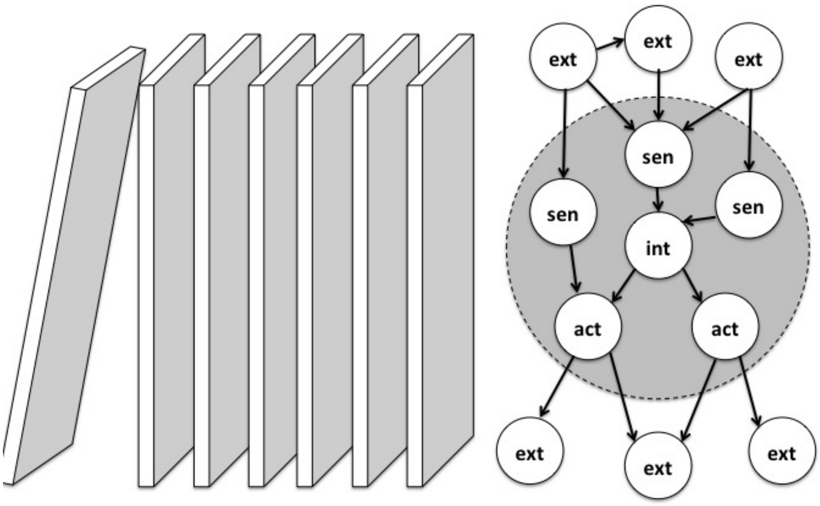


Figure 2.2: *Left* A domino chain reaction as an exemplificiation of a Markov process. Every domino is only dependent on its previous domino, just as, in the figure on the *right*, every node is only dependent on its neighboring nodes. *Right* A schematic depiction of a Markov blanket, a spatial generalization of a Markov process. The *gray circle* represents the Markov blanket of a node, consisting of internal state (int), its children (the action states, act), its parents (sen), and parents of children (sen), with parent/child being understood in terms of cause/effect.

dynamical system can be described naturally in terms of 'internal' and 'external' states 'shielded off' by a blanket of 'perception' and 'action states'. This allows for an inferential understanding of coupled dynamical systems. Friston claims that, over time, the intrinsic dynamics of the internal states (something like electrochemical dynamics) will start to 'infer' the causally disconnected states of the intrinsic dynamics of the environment (that are 'hidden' behind the Markov blanket).

Markov blankets introduce what Eliasmith (2005) calls a perspective of an animal on its environment. Do Markov blankets imply the inferential seclusion of the animal from its environment? We think this does not follow at all, and rather points to a direct coupling between animal and environment. Friston's simulation of the primordial soup shows that what he calls 'inference' is a natural consequence of processes governed by Newtonian and electrochemical interactions, completely understandable in terms of the basic and natural phenomenon of self-organization that Friston describes using the tools of graph theory. The notion of "inference" Friston is using is much more minimal than the one typically employed in philosophy, and refers to the mutual information (mutual predictability). Inference, for

Friston, is the process that leads to nodes inside of a Markov blanket and nodes outside of the blanket having high mutual information. We understand inference as implying a probabilistic relation between representational states. However, in the primordial soup simulation, we doubt that it makes sense to ascribe representational states to the systems that are being modeled. Instead, we can explain the process that leads to mutual information more parsimoniously in terms of synchronization.

The simulations in Friston (2013b) show that any dynamical system A coupled with another B can be said to "infer" the "hidden cause" of its "input" (the dynamics of B) when it reliably covaries with the dynamics of B and it is robust to the noise inherent in the coupling (Bruineberg & Rietveld, 2014 (Chapter 1), p. 7). We question whether it makes sense to call this inference, given that the process that leads to high mutual information need not be thought of as representational. Friston recognizes that the internal and external dynamics of an animal in its environment are coupled through something called generalized synchrony. Generalized synchrony is a well-known phenomenon in physics, stemming from Huygens (1673) study on the synchronization of clocks. It is worth taking a closer look at this phenomenon very briefly because it is the minimal basis for his notion of inference as these concepts are understood by Friston in his writings on the free-energy principle.[33]

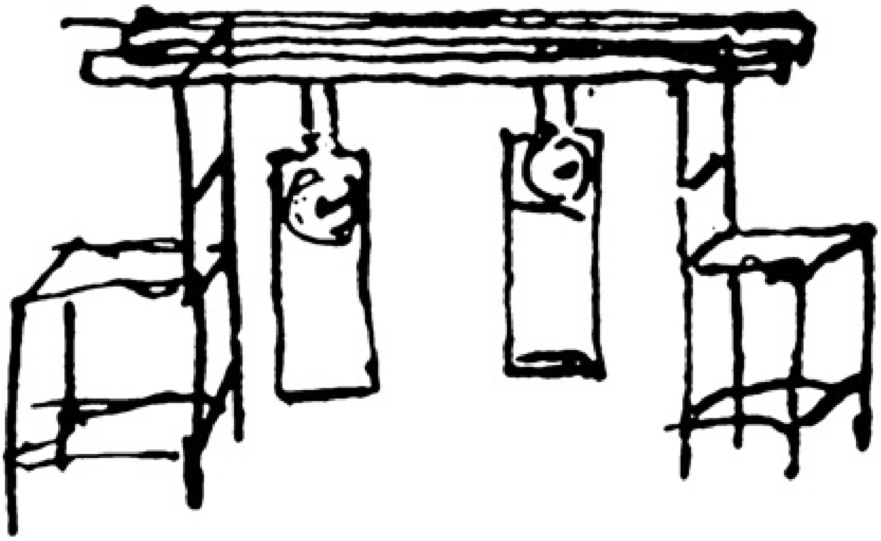


Figure 2.3: Original drawing from Huygens' C. Horoloqium Oscilatorium. Two oscillating clocks hanging on a suspension beam that is itself able to move

---

[33] We do not mean to argue that coupled clocks are just as much living systems as, for instance, the cell. As mentioned above in footnote 17, additonal conditions are required to demarcate living from non-living systems. Our point here is to highlight the scope of the concept of 'inference' that is at play in the free-energy principle. The concept of inference as Friston uses it applies to any process that leads to states with mutual information (**Figure 2.3**).

Huygens observed that two pendulum clocks in the same housing start to synchronize through the tiny movement of the beam on which they are suspended, or the table on which they are placed. Such synchronization is a general and common phenomenon in non-linear dynamics (Strogatz, 1994). When different time-scales are involved, one often speaks of 'enslavement', while when both systems exhibit similar typical times-scales one speaks of 'entrainment'. One can understand these dynamics in terms of a synergetic 'circular causality' (Tschacher & Haken, 2007). In systems with 'circular causality', there is no clear difference between internal dynamics attuning to external dynamics and vice versa (or clock 1 attuning to clock 2 and vice versa), rather both systems are bidirectionally coupled and reduce the disattunement between them until equilibrium (synchronization) is reached.

In this example the suspension beam, or table, acts as a Markov blanket. That is to say, the two clocks are statistically independent conditioned upon the movement of the beam. This is clear from the fact that no synchronization occurs when the beam is held fixed.[34] When, for instance, the intrinsic dynamics of the two clocks are too different, no synchronization will occur.

Based on Friston (2013b), one would be required to give an inferential interpretation of the coupling of the two clocks, in which one clock "infers" the state of the other clock hidden behind the veil of the connecting beam. In this example, each clock is a 'generative model' of the dynamics of the environment (the other clock), coupled through the Markov blanket of the connecting beam. The synchronizing clocks are therefore excellent examples of *being a model* (discussed in the previous section). Clock A is a model of clock B if clock A resonates with clock B. Whether synchronization (or "inference") occurs or not, is dependent on properties that have to do with the whole system: the period of the two clocks should not be too dissimilar; the beam should not be too rigid and not too flexible, and should be flexible in the right direction etc. We think such an inferential interpretation is unnecessary: there is no special inferential system inside of the Markov blanket, but the synchronization *is* the process of achieving high mutual information. This is emergent from the critically balanced coupled dynamics of the whole system.

Our point here is that, at least at this minimal level, there is no privileged explanatory role for the system that is defined 'within' the Markov blanket in achieving high mutual information with states of the environment. In the case of the two clocks, there is a perfect symmetry between 'internal' and 'external' dynamics. In the 'primordial soup' example of Friston (2013b), both the 'organism' and the 'environment' consist of essentially similar subsystems with electrochemical internal dynamics. It is only when these subsystems are similar enough that they will appear to exhibit active inference (i.e., to synchronize). Again, what

---

[34] Holding the beam fixed is equivalent to 'lesioning' the Markov blanket in Friston (2013b, Figure 5).

counts as similar enough is dependent on the system as a whole.[35] Given that attunement or synchronization takes place and given the 'environment', an observer can make predictions about the structure of the 'agent'. In line with the free-energy principle, one can show that the average information-theoretic surprisal of the coupling is minimized when the dynamics of coupled systems are synchronized. In this sense, synchronization is a form of free-energy minimization, which can be given a Bayesian interpretation (the posterior divergence goes to 0 when the clocks synchronize). By synchronizing, clock 1 can be interpreted as maximising its Bayesian model evidence for the state of the environment. Although one can describe the behaviour of these systems in terms of probabilistic inference, this is unnecessary. The process of achieving high mutual information can better be understood in terms of the coupled dynamics of the system as a whole.

One might object that there is still a non-trivial boundary separating the system from its environment in both the examples discussed above. Both systems are causally interacting through a Markov blanket. We agree, but we do not think this has any implications that conflict with the arguments of this paper. The importance of such a boundary for living organisms has been central in the autopoietic approach from the very start (Maturana & Varela, 1980; Varela et al., 1991). The mathematical methods used in ecological psychology also rely heavily on synergetics, synchronization and entrainment (Stepp & Turvey, 2010; Stepp et al., 2011; Haken et al., 1985), all presupposing *coupled* (rather than fused) dynamical systems. If this is the only kind of boundary that stems from the free-energy principle, then there seems to be nothing in the idea of probabilistic inference *per se* that challenges enactive cognitive science (Bruineberg & Rietveld, 2014 (Chapter 1); Clark, 2015b). In other words, *if* the role played by the concept of inference in FEP is no more cognitively demanding than generalized synchrony, *then* it provides no grounds for distinguishing radical enactive approaches to cognition from those that are based on FEP.[36]

One might object that there is an important difference between simple life-forms or primordial soup simulations and more complex systems that predictive-coding is seeking to model. The behavior of the first class of systems might be understood in terms of dynamic coupling and reducing disattunement. The second class of systems necessitates the explicit computation of prediction-error based on a generative model and requires a more epistemic understanding of

---

[35] An important difference between the clocks and the 'agent'-environment relation is of course that the agent and environment have different dynamical properties. Although conceptually more clarifying, the phenomenon of synchronization is not limited to systems with identical dynamics, but also applies to coupled non-identical systems (see for instance Pyragas, 1997).

[36] Importantly, generalized synchrony is not just used to model self-organizing systems. There is a body of literature that relates generalized synchronization to active inference more generally (see for example Friston, Sengupta, & Auletta, 2014; Friston, Levin, Sengupta, & Pezzulo, 2015; Friston & Frith, 2015).

the internal workings of the animal: wherever there are systems that can select actions so as to deal with the absent or lie in the future, this requires explicit representational knowledge structures about these distal affairs, one might think (Seth, 2013, Clark, personal communication).

However, as Orlandi (2016) points out as well, it is not clear that the structure of the generative model should be thought of as representational. We follow ecological psychology in arguing that the environment is rich with information that the perceptual system is able to directly access, because of our evolutionary history and the abilities and skills we learn. People and other animals (for instance primates) are situated in a very rich landscape of affordances (Rietveld & Kiverstein, 2014). We proposed that the generative model is best thought of as a dynamical system of (affordance related) states of action-readiness that reflect[37] the hierarchical and temporal organization of the changing environment. As the animal develops skills, the generative model becomes more and more sensitive to the relevant particularities of the situation, and opens the animal up to the relevant affordances available in the environment.[38]

Some of these abilities and skills for engaging with the world may be reused in ways that may explain our capacities for planning, imagining and reasoning about counterfactuals (Anderson, 2014). Other abilities for planning and imagining might come from the ways in which we actively structure the environment to scaffold our thinking (Rietveld & Kiverstein, 2014). The exact balance of environmental and agential components is an open empirical question to be settled on a case-by-case basis (see the ethnographic description of architects at work in Rietveld and Brouwers (2017)). However, we believe that the richness of the landscape of affordances in which human embodied minds are situated will be an important part of the answer.

We wish to make three final points before we conclude. First, the explicit computation of a prediction-error will be based on structures that represent and encode long term information about causal regularities in the environment. We have argued that for many cases of free-energy minmization, it is unnecessary to make appeal to this type of structures within the organism. We leave it as an open question when, and if at all, it is necessary to explain the internal structure of the organism by appeal to representations. For our purposes we prefer to think in terms of affordance-related states of action-readiness also for understanding forms

---

[37] By ‘reflect’, here, we mean that there are internal states of the organism that reliably covary with aspects of the environment, as discussed above in our analysis of the primoridal soup simulation.

[38] Anderson and Chemero (2013) raise the important worry that appealing to the concept generative models in a theory of the anticipating brain may lead one down ‘the lonely and unproductive road’ of epistemic internalism (p. 25). We think our interpretation of the generative model avoids this risk because our interpretation is non-representational. We understand the generative model in terms of multiple simultaneous states of action-readiness. These states of action-readiness are selectively sensitive to affordances available in landscape of affordances, which is there independently of any particular individual.

of what is traditionally called ‘higher’ cognition. Second, the free-energy principle provides a clear functional continuity between simple living systems and complex ones like humans. Predictive-coding theories understood in abstraction from the free-energy principle run the risk of missing this important functional continuity, understood by us in terms of the tendency towards an optimal grip. Third, one of the big questions for future work in neuroscience is how more locally mechanistic accounts of brain functioning, like predictive-coding, can be integrated with a more broadly complex systems perspective on the brain. Phenomena like entrainment and synchronization seem to play a crucial role in brain functioning (Varela, Lachaux, Rodriguez, & Martinerie, 2001; Engel, Fries, & Singer, 2001). We think the “ecological-enactive” reading of the free-energy principle provides a necessary framework for carrying out this integrative work. This is something one would fail to accomplish if one would exclusively focus on predictive-coding and ignore the free-energy principle.

2.5 Conclusion

The Helmholtzian interpretation of the anticipating brain brings with it a problematic conceptualization of the boundary separating the organism from the environment. The environment becomes a hidden cause that must be inferred from ambiguous sensory input, which is all the brain has access to. However we have argued above that there is no reason to accept such a Helmholtzian interpretation of the relation between the organism and its environment. Moreover, such an interpretation is in deep tension with the free-energy principle according to which the brain is a part of a larger coupled system that on the basis of its coupling is constantly reducing disattunement with the environment. Through the organism’s minimization of free-energy, the brain’s internal dynamics are normally adequately attuned to the external dynamics of the environment.

Central to the predictive processing framework is the notion of a generative model. On the Helmholtzian interpretation, a generative model is a model that captures the causal structure of the agent’s environment. The generative model is then used to generate ‘mock’ sensory input that is then compared to actual sensory input to compute prediction-errors. Such an interpretation introduces a functional asymmetry between (top-down) predictions and (bottom-up) prediction-errors. Top-down predictions are produced on the basis of internally realized knowledge structures (Clark, 2016a). This may raise the worry that our ecological-enactive interpretation fails to accommodate this functional asymmetry.[39] Without these internally realized knowledge structures, we just have complex looping interactions between brain, body and environment that mysteriously reduces disattunement.

On our interpretation, the function of the generative model is not to provide the agent with a representation of the dynamical structure of the environment *per se*, but rather to steer its interactions (over multiple timescales) with its environment in such a way that a *robust* brain-body-environment system is maintained. The organism's internal structure and organization is then understood as multiple simultaneous and coupled affordance-related states of action-readiness that together shape (through top-down precision-modulation) the salience of solicitations in the environment. The self-organization of these states of action-readiness allows the animal to tend towards an optimal grip on the multiple relevant affordances in the situation. There is an asymmetry between the sensory inputs that flow from the environment and the anticipations of sensory input based on the organism's history of interactions with the environment. However, we have been arguing that the internal dynamics can be understood in terms of patterns of action-readiness that constitute selective openness to multiple affordances simultaneously rather than from internalized knowledge structures. As we have proposed in earlier work (Bruineberg & Rietveld, 2014 (Chapter 1)), the tendency towards an optimal grip on multiple affordances simultaneously can be explained as a metastable attunement to environmental dynamics. This metastable attunement allows for rapid and flexible switching between relevant action possibilities (Kelso, 2012; Rietveld, 2008b; Bruineberg & Rietveld, 2014 (Chapter 1)). Crucially, once we dispense with misleading/distracting talk of probabilistic inference, it is no longer necessary to understand past experience and learning as encoded in the form of representational knowledge structures. Instead we understand past learning and experience as manifest in the skilled animal's anticipatory dynamics to act in ways that improve grip on the affordances on offer in the situation.

Grip, as so conceived, is certainly not "the new representation" as Andy Clark has suggested (Clark, 2015b). On Clark's view, the generative model is estimating the current state of the organism in its environment, something that is only possible because of the knowledge the agent has of the world. On our view of the generative model, it is preparing the agent to perform actions that improve grip on multiple simultaneously relevant affordances in the situation. This is something that is only possible because of metastable attunement to environmental dynamics that make the agent open and ready for dealing effectively and flexibly with what the future may offer.

---

[39] Andy Clark raised this objection to us in personal communication and on http://philosophyofbrains.com/2015/12/15/conservative-versus-radical-predictive-processing.aspx.

Chapter 3

Active inference and the primacy of the ‘I can’[1]

Abstract

This paper deals with the question of agency and intentionality in the context of the free-energy principle. The free-energy principle is a system-theoretic framework for understanding living self-organizing systems and how they relate to their environments. I will first sketch the main philosophical positions in the literature: a rationalist Helmholtzian interpretation (Hohwy, 2013; Clark, 2013), a cybernetic interpretation (Seth, 2014) and the enactive affordance-based interpretation (Bruineberg & Rietveld, 2014 (Chapter 1); Bruineberg, Kiverstein, & Rietveld, 2016 (Chapter 2)) and will then show how agency and intentionality are construed differently on these different philosophical interpretations. I will then argue that a purely Helmholtzian is limited, in that it can account only account for agency in the context of perceptual inference. The cybernetic account cannot give a full account of action, since purposiveness is accounted for only to the extent that it pertains to the control of homeostatic essential variables. I will then argue that the enactive affordance-based account attempts to provide broader account of purposive action without presupposing goals and intentions coming from outside of the theory. In the second part of the paper, I will discuss how each of these three interpretations conceives of the sense agency and intentionality in different ways.

3.1 Introduction

After computationalism, connectionism, and (embodied) dynamicism, cognitive science has over the last few years seen the resurgence of a paradigm that might be dubbed “predictivism”: the idea that brains are fundamentally in the business of

---

[1] This paper was previously published as Bruineberg (2017).

predicting sensory input. This paradigm is based on older ideas in psychology and physiology (Helmholtz, 1867), and has been revived by parallels that have been discovered between machine learning algorithms and the anatomy of the brain (Dayan & Hinton, 1996; Friston et al., 2006). The emergence of the paradigm of "predictivism" has sparked great interest in philosophy of mind and philosophy of cognitive science, mainly through the work of Clark (2013, 2016b) and that of Hohwy (2013, 2014). This interest has led to a vast number of papers attempting to ground concepts from phenomenology, philosophy of mind and psychopathology in predictive architectures (see for example Hohwy, 2007; Limanowski & Blankenburg, 2013; Apps & Tsakiris, 2014; Hohwy, Paton, & Palmer, 2015).

Predictivism might be better off than these earlier paradigms in cognitive science, exactly because most of its core ideas are *not* very new. As Clark writes in the introduction to his book:

> [W]hat emerges is really just a meeting point for the best of many previous approaches, combining elements from work in connectionism and artificial neural networks, contemporary cognitive and computational neuroscience, Bayesian approaches to dealing with evidence and uncertainty, robotics, self-organization, and the study of the embodied environmentally situated mind. (Clark, 2016b, p. 10)

To put it in Kuhnian (1962) terms, for Clark we might currently see the transition of cognitive science from a pre-paradigmatic stage, with competing paradigms developed by incompatible schools of thought, to *normal science* in which one dominant paradigm provides the concepts and questions to be solved. Whether or not this is true is for Kuhn a question that can only be answered in hindsight. In any case, by providing a meeting point for these different approaches, "predictivism" simultaneously also provides a new battleground for competing schools of thought in philosophy of mind concerning internalism and externalism, embodiment, and computationalism.

Currently, it is unclear whether "predictivism" entails a particular philosophical position, and whether "predictivism" tells us much about the nature of cognition without these philosophical assumptions frontloaded. Different scientists and philosophers working on predictive-coding take different, supposedly mutually incompatible starting points: a Helmholtzian theory of perception (Hohwy, 2013; Clark, 2013), Ashbyian cybernetics (Seth, 2014) and an enactive affordance-based account borrowing from Merleau-Ponty and Gibson (Bruineberg & Rietveld, 2014 (Chapter 1); Bruineberg et al., 2016 (Chapter 2); Rietveld et al., in press).[2] To me, there seems to be little hope to settle philosophical issues concerning embodiment and the mind-world relationship deriving from a theory-neutral presentation

---

[2] I do not wish to say that these positions are *a priori* mutually exclusive. However, they do have very different philosophical starting points and it therefore remains to be seen to what extent they are (in)compatible.

of predictive processing (PP). In fact, as mentioned in the introductory chapter (Wiese & Metzinger, 2017) a theory-neutral presentation of PP seems itself unfeasible. Rather, much of the literature poses a problem in which a philosophical worldview is presupposed and then shows the compatibility of PP with this view, be it about using sensory input to represent a distal world (Hohwy, 2014, p. 1), tending towards grip on a field of affordances (Bruineberg & Rietveld, 2014 (Chapter 1), p. 7), or the problem of homeostatic regulation and interoceptive inference (Seth, 2014).

In this paper, I will focus on how to conceive of agency and the sense of agency under the free-energy principle (FEP). The free-energy principle is the most theoretical and all-encompassing version of the "predictivist" approach, being compatible with, but not limited to, predictive-coding accounts of the brain. In itself, the free-energy principle is a system-theoretic framework for understanding living self-organizing systems and how they relate to their environments. I will first present the main tenets of the free-energy principle and consequently present three different philosophical approaches to the free-energy principle: a rationalist approach (based on Helmholtz), a cybernetic approach (based on Ashby) and an enactive affordance-based approach (based on Merleau-Ponty and Gibson). I will argue that whereas the rationalist and cybernetic approaches face a number of conceptual problems in construing agency under the free-energy principle, these conceptual problems can be resolved by the enactive affordance-based approach.

3.2 Main tenets of the free-energy principle

In this section I will give a non-mathematical treatment of the basic tenets of the free-energy principle, introducing the main assumptions and reasoning steps that lead to its formulation. [For an introduction to predictive processing and the free-energy priniciple more generally, see Wiese and Metzinger (2017) and references therein.]

As mentioned in the introduction of this paper, the free-energy principle is a proposal for understanding living self-organizing systems (Friston & Stephan, 2007; Friston, 2011). Based on a descriptive statement (living systems survive over prolonged periods of time), the free-energy principle provides a prescriptive statement (a living system must minimize its free-energy) to provide the necessary and sufficient conditions for this descriptive statement to be true. The major premises underlying this move are the following:

1. The embodiment of an animal implies a set of viable states of the animal-environment system.

One can formalize this in information-theoretic terms by assigning a probability distribution to the viable states of the organism. For example, human body temperature has a high probability of being around 37° and a low probability of being elsewhere. Information theoretically, this means that the event 'measuring a body temperature of 37°? has low surprisal, while measuring a body temperature of 10° has a very high surprisal. Remaining within viable bounds can then be understood in terms of minimizing surprisal. For ectothermic (cold-blooded) animals, this directly puts constraints on the places in its environment that it may seek out (i.e., a lizard seeking out a sunny rock in the morning). For endothermic (warm-blooded) animals, this means it needs to find energy sources to sustain its metabolism and, in some cases, seek shelter to complement its internal heat regulation. In short, with a particular agent we can identify a probability distribution of the states the agent typical frequents and has to frequent. I will call this distribution the *embodied* distribution and the surprisal of an event relative to this *embodied* distribution embodied suprisal (see Bruineberg et al. (2016 (Chapter 2)) for a more elaborate introduction of this vocabulary and see Wiese and Metzinger (2017) for an informal analysis of how this distribution can be found based on the typical states the animal frequents and the assumption of ergodicity).

2. The animal's regulatory system (for instance the nervous system) does not have access to the viable states of the agent-environment system. Instead it needs to estimate them.

A regulatory system needs to minimize surprisal without being able to evaluate it directly. It cannot evaluate surprisal directly, because the embodied probability distribution of the viable states of the organism is not known to it. This is where free-energy comes in. Free-energy is a function of sensory states and estimated worldly states that generated the sensory states and involves two probability densities:

- A generative density $p(w, s|m)$, specifying the joint probability of sensory state s, and worldly states w based upon a probabilistic model m embodied by the agent.
- A recognition or variational density $q(w; b)$, encoding the agent's 'beliefs? about the worldly states w entailed by its internal state b.

Free-energy is defined in terms of these two densities:

$$F(s, b) = -\int_w q(w, b) \ln \frac{p(w, s|m}{q(w; b)} dw \tag{3.1}$$

The free-energy formulation can be rearranged so as to show its dependence on perception and action respectively (see Friston & Stephan, 2007; McGregor, Baltieri, & Buckley, 2015). The basic idea behind the free-energy framework is that whatever shape or form the recognition density takes, free-energy over the long run related to this *estimated* recognition density will be equal to or greater than the surprisal I receive at any point in time related to the *embodied* distribution. The long-term average of free-energy (obtained by integrating over the temporal domain) is called free action.

The quantity of free-energy is a function of sensory states and estimated worldly states and priors. Each of these can change in order to minimize free-energy: optimizing estimated worldly states (typically called perceptual inference), optimizing sensory states (brought about through action), and optimizing the generative model (learning).

3. In order to stay alive, it suffices for the animal to stay within the viable states of the animal-environment system. It does so by minimizing free-energy using its estimated conditions of viability as priors.

The assumption here is that the internally estimated conditions of viability and the real (embodied or intrinsic) distribution are similar enough to make adequate regulation possible (i.e., my regulatory system should not anticipate a body temperature of 10°). Homeostatic control can then be achieved by predicting particular sensations corresponding to a body temperature of around 37° and minimizing the discrepancy from, or prediction-error with respect to that implicit hypothesis (the expectation of body temperature of around 37°). The logic here is that through evolution and development the agent comes to expect itself to be in an optimal state and continually minimizes the discrepancy between its current state and its optimal expected state.

4. To achieve homeostatic control, the animal needs to be able to act on the world. This implies, at least implicitly, a model of how actions lead to changes in interoceptive and exteroceptive sensory input. Since the state of the world mediates how actions change perception, optimal regulation requires taking the state of the world into account. This process of optimal regulation while taking the estimated state of the world into account is called 'active inference'.

As we will see later in this paper, perceptual inference will only minimize free-energy to the extent that it becomes a tighter upper bound on embodied surprisal: the animal becomes more and more certain of it being in a too cold states. Only action will change embodied surprisal (the animal being in a too cold state) itself. For example, the ectothermic lizard needs to be able to compensate its interoceptive prediction-error by moving around in a way that makes it seek out the warm rock in the sunshine.

To summarize: to continue being a living creature is to maintain oneself in a particular type of environment. For example, to be a fish the fish must maintain itself in a fish-like environment. This is possible for the fish through its prediction of the sensory input associated with a fish-like environment (a certain pressure, temperature, light etc.) and through its actions (being able to avoid, accommodate and counteract mismatches between predicted sensory input and actual sensory input). This implies a particular kind of congruence between the dynamics and structure of the environment and of the organism, to which I will return in a later section.

3.2.1 Prediction-error minimization and the free-energy principle

The free-energy principle does not provide in and of itself a mechanism for realizing free-energy minimization. However, it often gets paired, or even conflated, with the prediction-error minimization framework (PEM).[3] PEM can best be introduced as a form of Bayesian model-based statistical inference. The animal possesses an internal model of the possible causal structures of the world and the kind of sensory information associated with these causal structures. Based on its priors and sensory states, weighted by its confidence in both, it can then infer the hidden state of the environment based on a series of sensory states. Adequate inference and adequate prediction are then two sides of the same coin.

Much work in computational neuroscience and machine learning has been carried out in the PEM framework with the aim of understanding how inference through prediction-error minimization is possible in brains. One important feature is that the generative model is hierarchical: each layer of the hierarchy tries to predict the information it is receiving from a lower level (Friston, 2008). Another central feature of this work concerns whether the agent's probability distribution is updated so as to approximate Bayes' theorem as the agent is exposed to new sensory input, and if so, which approximation algorithms work best. These developments might make predictive-coding neural architectures a good computational implementation for free-energy minimization.

However, there remain a number of conceptual tensions between machine learning approaches and the free-energy principle that will be the main focus of my discussion in the remainder this paper. They concern the role of action: is action auxiliary in obtaining the most likely hypothesis or is action goal-directed? If there is no strong distinction between the two, how can we conceive of both the epistemic and the goal-directed function of action simultaneously? In the next section, we will see how different philosophical approaches respond to these questions.

---

[3] For example, Hohwy writes: "Since the sum of prediction error over time is also known as free-energy, PEM is also known as the free-energy principle" (Hohwy, 2014, p. 2).

3.3 Philosophical interpretations of predictivism

In this section, I will present three philosophical approaches to the free-energy principle: a Helmholtzian approach (Hohwy, 2013; Clark, 2013), a cybernetic approach (Seth, 2014) and an enactive affordance-based account borrowing from Merleau-Ponty and Gibson (Bruineberg & Rietveld, 2014 (Chapter 1); Bruineberg et al., 2016 (Chapter 2)). I will discuss how they conceive of action under the free-energy principle.

3.3.1 Helmholtz and hypothesis-testing

The standard point of departure for "predictivist" approaches to the mind is Helmholtz's (Helmholtz, 1867) notion of perception as unconscious inference (more recently, Gregory (1980) articulated the idea of perception as hypothesis-testing). The basic idea is that perception is essentially continuous with the scientific method. That is to say, the perceptual system holds a hypothesis (or a range of hypotheses) with a certain degree of confidence. Incoming data might corroborate the current hypothesis, cause the system to change its hypothesis, or cause it to abandon it altogether (i.e., shift to a new hypothesis). By iterating this process over time, the system comes to infer the true hidden state of the environment. Expected precision (i.e., degree of confidence) of both the hypothesis and the sensory input plays a crucial role in how (and whether) one settles on a particular hypothesis. For example, in a perceptual decision-making trial, I might start out in a very low confidence state. Over time, while sensory information comes in, I develop a hypothesis (say, dots on average moving to the right) that explains away the prediction-error. Over time, the confidence in the hypothesis grows until a threshold is reached. Noise in the system has a high impact in the beginning of the trial, when confidence in the current hypothesis is low and has low impact in the end, when there is high confidence (see Bitzer, Bruineberg, and Kiebel (2015) for an example of such a Bayesian model of perceptual decision-making).

The Helmholtzian perspective seems to work well for perceptual inference. For Helmholtz, as for Gregory, it is *perception* that is inferential. They implicitly endorse a 'sandwich model of the mind' (Hurley, 1998): perception supplies input to the cognitive systems, which figure out what to do next, and action translates decisions into motor commands. This is not to say that Helmholtzians think perception is "dumb" or passive. Contrary to other sandwich models (Fodor, 1983), Helmholtzian perception is thought to be active and knowledge based. Modern "predictivist" accounts depart from Helmholtz in the sense that they deny the sandwich model, and attempt to closely intertwine perception and action in what is called "active inference". Regardless of these differences, I take the basic commitments of Helmholtzian cognitive science to be 1.) That the *aim* of perception and action is to disambiguate the hidden causal structure of the environment, and 2.) That the *means* by which this aim is achieved is by some process continuous with or analagous to scientific hypothesis-testing

At first glance, active inference might only seem to strengthen the link between the workings of the mind and scientific inferences. The way Friston (2012) and Hohwy (2013) add action to perceptual inference is by appealing to setting up experiments. Scientific hypothesis testing is not just passively recording results of data coming in, but carefully setting up experiments and actively intervening in the chains of causes and effects in order to disambiguate the hidden causes of sensory input. Hohwy relates this to causal inference (Pearl, 1988) in which the system is able to calculate where to intervene in order to disambiguate between causal structures. This is an elegant way of combining perception and action under the umbrella of "predictivism".

However, the only demand on a perceptual system is whether it is adequately able to infer, represent and predict the hidden state of the environment. It lacks an account of motivation, value and reward—on whether a particular environmental state is conducive or detrimental with respect to its bodily needs, habits or plans. As we have seen in **Section 3.2**, the free-energy principle *does* aim for such deeper integration of prediction and motivation (Friston et al., 2009). The way in which it does so is by reducing the traditional roles of cost functions, value and reward to prediction. However, I will now argue that, in doing so, it fundamentally changes the very nature of prediction error-minimization.

Consider the following example as an intuition pump: suppose I am standing under a steaming hot shower. This will lead to prediction-errors on the skin. There may be some physiological reactions that might help to reduce temperature (such as vasodilation), but the most obvious reaction would be to get out of the shower, or to manually change the temperature. This requires an implicit generative model of how interoceptive and exteroceptive prediction-errors change with particular actions and not others, while taking into account the peculiarities of the shower I am standing under now.

What is important here is that the most *likely* cause of the sensory input I am receiving is the fact that "I am standing under shower that is too hot" and any "experiment" I set up will corroborate that hypothesis. What the system *needs* to do is to treat burning under a hot shower as *extremely unlikely.* Since it is extremely unlikely, I cannot accept this hypothesis, but rather am forced to change the world so as to reduce prediction-errors with respect to the hypothesis that "I am standing under a comfortably warm shower". To emphasize: although FEP treats the current state as highly unlikely, it *is* the actual state I find myself in and if I do nothing I *will* get burnt. *If* the aim of the Helmholtzian account is solely to figure out what the hidden state of the world is, then "I am standing under a shower that is too hot and will get burned" will be the hypothesis it settles for, but it is not. This gives rise to the *crooked scientist argument*: if one wishes to compare the activities of the brain to that of a scientist, it *needs* to be a ‘crooked scientist’. The brain acts like a scientist invested in ensuring the truth of a particular theory, which is the theory that “I am alive”.[4] As contradictory evidence comes in, it manipulates the world until the perpetual truth of that theory is ensured (or dies trying) (Bruineberg et al., 2016 (Chapter 2)).

I believe there is an important shift here in the conceptualization of active inference. On the Helmholtzian picture, a system does better when it is more accurate and precise in its representations of the causal structure of the environment, i.e., when it delivers true representations that the perceiver has high confidence in. On the ecological enactive picture that I will sketch, a system performs better when it supports the system’s movement towards an optimal state, where that optimal state is to be understood relative to the animal’s conditions of viability/flourishing. In that sense, active inference requires a thoroughly optimistic generative model of how the animal expects to flourish in its econiche. Only with such an optimistic generative model will active inference lead to adaptive behavior.

Note that within this picture there is ample room for epistemic actions like those the scientist performs in carefully setting up an experiment. Consider the everyday example of standing in a small space under a too hot shower while having shampoo in your eyes. One can imagine the following response to the situation: I first orient myself, for instance by touching the wall, and then reach for the tap to turn down the temperature. The first action can be seen as largely epistemic, the second one as largely goal-directed.[5] However, epistemic actions unfold within the context of the movement towards an optimal state, where the optimality, in this context, is grounded in the system’s conditions of viability/flourishing.

Hohwy also seems to want to ground or justify prediction-error minimization by appealing to biological self-organization. He raises the issue in the context of Kripke’s (1982) interpretation of Wittgenstein’s (1953) rule following argument. The skeptical question, phrased in predictive-coding terms, is whether it makes sense to say of someone that he or she correctly obeys the imperative of minimizing prediction-error. What is the fact of the matter about that person that justifies the assertion that he or she is correctly obeying that rule? Hohwy states: “The answer cannot be something along the lines of: you should minimize prediction-error because it leads to truthful representations. That answer is couched in semantic terms of ‘true representations’, so it is circular” (Hohwy, 2013, p. 180). He finds his way out of this circularity by appealing to a non-semantic feature of

---

[4] Of course, staying alive underdetermines what to do in everyday situations. For such cases, enacting a (more or less coherent) identity, “flourishing” or having grip on the situation might be better suited notions.

[5] In a recent paper, Friston, Rigoli, et al. (2015) show that one can mathematically decompose free-energy minimization into epistemic value and extrinsic (goal-directed) value (Seth (2015) calls these epistemic and instrumental, respectively). Epistemic value serves to reduce uncertainty related to hidden states of the world (i.e., the location of the tap), while extrinsic value serves to bring the agent closer to an optimal state.

our existence: self-organization. "I should minimize prediction error because if I do not do so then I fail to insulate myself optimally from entropic disorder and the phase shift (that is, heat death) that will eventually come with entropy" (Hohwy, 2013, p. 181). I agree with Hohwy that any notion of predictive-processing needs, in order to avoid being circular, ultimately to be grounded in the requirement for biological self-organization. But I disagree on the constraints this requirement places on how to conceptualize PEM. Hohwy continues:

> Perhaps we can put it like this: misrepresentation is the default or equilibrium state, and the fact that I exist is the fact that I follow the rule for perception, but it would be wrong to say that I follow rules in order to exist? I am my own existence proof. (Hohwy, 2013, p. 181, my italics)

If by 'the rule of perception' Hohwy means that our perceptual system is in the business of minimizing prediction-error (and action is at most auxiliary), then we are in disagreement about what grounding PEM in biological self-organization entails for PEM. As the crooked scientist argument shows, accepting the requirements for biological self-organization entails a shift from tending towards a more truthful representation to tending towards a more optimal agent-relative equilibrium. It is exactly this shift, and its implications for how to conceptualize agency, that remain concealed on the Helmholtzian account.

As mentioned above, I take the basic commitments of the Helmholtzian cognitive scientist to be 1.) that the *aim* of perception and action is to disambiguate the hidden causal structure of the environment, and 2.) that the *means* by which this *aim* is achieved is by some process continuous with scientific hypothesis-testing. Based on Friston, Rigoli, et al. (2015) and the crooked scientist argument, I take it that commitment 1 is false in a strict sense. At best, figuring out the hidden structure of the world is *auxiliary* to moving towards a more optimal state. I take commitment 2 to be false in a strict sense as well. In Helmholtzian language, perception and action serve to optimize the likelihood of the animal's theory that it is alive. If certain "experiments" don't give the right answer, the animal will switch to performing new "experiments" that do give the right answer: I change the temperature of the shower and not the hypothesis about the kind of being that I am (i.e., one that survives at 37°).

The 'crooked scientist argument' is problematic for those who wish to endorse both the free-energy principle and a Helmholtzian theory of cognition. The Helmholtzian metaphor gives you exactly the *wrong* intuitions about some core aspects of the free-energy principle. The intuition in active inference should not be, as Hohwy claims, how the brain can use available sensory input to accurately reconstruct the hidden state of affairs in the world (Hohwy, 2014, p. 1), but rather how the space of possible 'hypotheses' is always already constrained and crooked in such a way as to make the animal tend to optimal conditions.

It is the analysis of (organism-relative) value as prediction error that makes the free-energy principle such a challenging framework to understand. Appealing to Helmholtzian inference does very little to make these conceptual difficulties clearer. As I hope to have shown, particular aspects of the Helmholtzian framework might be retained, but overall it does a poor job. Furthermore, I think that both the "cybernetic Bayesian brain" (Seth, 2014) and Merleau-Pontian cognitive science (in the form of the "Skilled Intentionality Framework" (Bruineberg & Rietveld, 2014 (Chapter 1); Rietveld et al., in press) and "Radical Predictive Processing" (Clark, 2015b; cf. also Downey, 2017) might be better alternatives. I will turn to them next.

3.3.2 Ashby and cybernetics

Rather than starting from Helmholtz, Anil Seth (2014) takes the work of cyberneticist Ashby (Ashby, 1952, 1956) as his starting for theorizing about "predictivism". Cybernetics focuses on control systems. The field is dubbed cybernetics after its prototype: the Watt governor (*κυβϵρνήτης* is Greek for governor), a clever device capable of stabilizing the output of a steam engine based on a system of rotating flyballs that controls the throttle valve (e.g., Van Gelder, 1995; Bechtel, 1998). The point about the Watt governor is that it is able to suppress perturbations in the system and, in doing so, stabilizes the governor-engine system. Inspired by the governor, cybernetics proposes the more general principle that an adaptive system maintains its own organization by suppressing and responding to environmental perturbations. This often includes the control of so-called *essential variables*. For example, in the case of living systems, body temperature and metabolic needs – when action and perception are coupled with a temperature sensor, a system might move through a space in order to seek out a place where the temperature is optimal.

The basic principles of cybernetics seem to fit well with the basic tenets of the free-energy principle: active homeostatic control to stay within viable bounds. Auletta (2013) provides a nice example of how a coupled informational (sensorimotor) and metabolic system can provide a model for bacterial chemotaxis and how this can be understood in terms of free-energy minimization. In simple and stable environments, such as are often used in evolutionary robotics, it is sufficient to train a simple neural network to pick up stable regularities between sensors, aspects of the environment and the availability of heat or food. In his work on embodied predictivism and embodied cognition more generally, Clark (1997, 2016b) has proposed that we understand the internal workings of the animal as such a "bag of tricks", fit to deal with its niche.

Seth's (2014) proposal makes progress in relation to one of the main weaknesses of the purely Helmholtzian account: the exclusion of values. Demanding that particular essential variables are kept constant puts constraints on the interactions the animal has with the environment. The example Seth gives is of that of blood sugar level. When blood sugar level is too low, the following responses arise: interoceptive prediction-errors signals travel upward in the brain, which lead to subjective experiences of hunger or thirst. These prediction-errors then travel further upward in the hierarchy where multimodal integration of interoceptive and exteroceptive inputs take place. These high-level models then instantiate predictions that flow down the hierarchy, leading to an autonomic response (metabolize bodily fat stores), or allostatic actions (eating a banana) (Seth, Suzuki, & Critchley, 2011; Seth, 2014). Contextual information, about for instance the availability of food (encoded in the precision of the allostatic response hypothesis), might contribute to the decision as to which response is initiated (or whether both are).

Similarly, the cybernetic account can handle the hot shower example given in the previous section. The hot shower will lead to prediction-errors (perhaps showing in the form of pain and dizziness) that stand in need of being reduced. This then puts constraints on the actions I might undertake, leading to the combination of epistemic and purposeful (extrinsic) actions that make me leave the shower or reduce the heat of the shower. In short, the cybernetic account is better suited to explaining adaptive and ecological action than the Helmholtzian account.

One other aspect of the free-energy principle that comes to the foreground on the cybernetic interpretation is the structure of the generative model. If the function of the generative model shifts from inferring the hidden state of the environment to steering the animal towards an optimal state, then the generative model is not just a model of the environment, but rather of the animal situated in its environment. What counts as the most likely state is not the most likely state of the environment *per se* given current sensory evidence and one's prior beliefs, but rather the optimal state for the animal-environment system to be in (Friston, 2011). I will return to this point in the next section.

On a purely cybernetic account, all actions are responses to (or responses of anticipations to) deviations from homeostatic variables. Seth's (2014) model of active inference integrates both cybernetic and Helmholtzian elements: action can both serve to confirm, disconfirm and disambiguate hypotheses as on the Helmholtzian account and also account for homeostatic behavior. Although the cybernetic account improves upon the Helmholtzian account, it is still limited in the account it gives of active inference. What is lacking is that the optimality conditions that the animal generates are broader than essential variables related to homeostasis. My metabolic needs underconstrain, for instance, in what way I will finish this sentence. Being an academic philosopher, the practices I participate in and the skills I have acquired in these practices, *do* constrain my writing style. Some of these practices and habits, like working and skipping dinner in order to finish a paper, might actually squarely oppose those metabolic needs. The challenge ahead, as I understand it, is to provide an account of non-metabolic purposes without an appeal to goals as unexplained explainers. The answer lies, I believe, in understanding how our purposive actions are situated in a social setting with which we are familiar. For these reasons, I will turn to a Merleau-Pontian approach to cognitive science next.

3.3.3 The enactive affordance-based account

A third philosophical perspective on "predictivism" can be distilled from the work of French phenomenologist Maurice Merleau-Ponty. The great insight put forth in Merleau-Ponty's *The Phenomenology of Perception* (Merleau-Ponty, 1945/1962) is that, as skilled humans, we have a pre-reflective bodily engagement with the world, prior to any objectification: we bike home from work, cook dinner and have a conversation. In such cases, we do not continuously decide what to do, but are open to and respond to the demands of the situation. According to Merleau-Ponty, a perceptual scene does not show up as a set of objects but is colored or structured by the demands of the situation:

> For the player in action the football field is not an 'object' [...]. It is pervaded with lines of force [...] and articulated in sectors (for example, the 'openings' between the adversaries) which call for a certain mode of action and which initiate and guide the action as if the player were unaware of it. [...]; the player becomes one with it and feels the direction of the 'goal', for example, just as immediately as the vertical and the horizontal planes of his own body. It would not be sufficient to say that consciousness inhabits this milieu. At this moment consciousness is nothing other than the dialectic of milieu and action. Each maneuver undertaken by the player modifies the character of the field and establishes in it new lines of force in which the action in turn unfolds and is accomplished, again altering the phenomenal field. (Merleau-Ponty, 1942/1966, pp. 168-169)

What we perceive in skilled action are the relevant action possibilities that the situation provides. We perceive these possibilities not as mere theoretical possibilities, but as what Dreyfus and Kelly (Dreyfus & Kelly, 2007) call relevant affordances or solicitations.

What is perceived as relevant depends on the situation, the skill of the agent and socio-material norms the agent is attuned to. Everything the football player has learned through years of practice feeds back in the way the situation appears. This tight coupling between skilled agent and environment, in which every action modifies the experiential field, is what Merleau-Ponty calls "the motor-intentional arc" (Merleau-Ponty, 1945/1962; Dreyfus, 2002).

There is a second notion borrowed from Merleau-Ponty that is important for our current purposes, and that is the notion of the "tendency towards an optimal grip". This is a primarily phenomenological notion that signifies the way a skilled individual relates to its environment. Merleau-Ponty gives the example of perceiving a picture in an art gallery: "There is an optimum distance from which it requires to be seen" (Merleau-Ponty, 1945/1962, p. 352). The details of the painting get lost when we step further away, and we lose the overview of the painting as a whole when we move too close. In a sense, the painting *demands* a particular perspective, just like the situation on the football field demands an action to be made. Note that, for Merleau-Ponty, absolute grip is never obtained, but it is the *tendency towards* grip that guides our actions.

A third insight from Merleau-Ponty that might be of help in the current context is the manner in which active agents bring forth their own world. Clark (2016b, p. 289), drawing upon the continuity between Varela et al. (1991) and Merleau-Ponty (1945/1962) writes:

> In a striking image, Merleau-Ponty then compares the active organism to a keyboard which moves itself around so as to offer different keys to the "in itself monotonous action of an external hammer" (Merleau-Ponty, 1945/1962, p. 31). The message that the world 'types onto the perceiver' is thus largely created (or so the image suggests) by the nature and action of the perceiver herself: the way she offers herself to the world. The upshot, according to Varela et al. (1991, p. 174) is that 'the organism and environment [are] bound together in reciprocal specification and selection'.

Now, as Clark is careful to note, the world is more than just a brute hammer, but the important message here is that the active agent meets the world on its own terms. This phenomenon is labelled differently in different traditions: ecological psychologists speak of perturbations being not *given by*, but *obtained from* the world (Turvey & Carello, 2012), autopoietic enactivists speak of an *autonomous* system bringing forth significance (Varela et al., 1991; Thompson, 2007; Di Paolo, 2005). The demand for self-organization provides, for both the free-energy principle and autopoietic enactivism, specific constraints on the circularity (sometimes called "circular causality", see Tschacher and Haken (2007)) between organism and environment: the environment and skilled agent mutually constrain each other in such a way that the overall dynamic remains within a flourishing regime.

3.3.4 Selective openness and active inference

In Bruineberg and Rietveld (2014 (Chapter 1)), we attempted to frontload Merleau-Ponty's notions of the intentional arc and the tendency towards an optimal grip within the free-energy principle in what is called the *Skilled Intentionality Framework* (see also Rietveld & Kiverstein, 2014; Rietveld et al., in press). The central tenet of active inference is that perception and action jointly minimize the discrepancy between actual and anticipated sensory input. However, as we have seen, the *goal* of active inference is not, as on the Helmholtzian account, to infer the hidden causes of the environment (at most, this is auxiliary), but rather to steer its interactions with the environment in such a way that a *robust* agent-environment systems is maintained in which the agent is flourishing. There is an intricate circularity built in at the heart of the free-energy principle: only when I predict myself to be an agent acting in the world, and flourishing in my environment, does minimizing prediction-errors lead to a flourishing state. Hence, if the agent's model is a generative model of something, it is a model of the agent acting in its niche (see Friston, 2011) and of how its own actions will change its exteroceptive and interoceptive sensations. I have suggested extending this circularity to include not just regulation of metabolic needs but also to incorporate attunement to the regular ways of acting (norms) of the patterned practices the agent participates in. For example, the way an agent responds to an outstretched hand has, arguably, no bearing on her viability conditions (as long as physical harm is avoided), but refusing to shake someone's hand might be seen as a violation of a social norm or as a political statement.

To return to the earlier example: the expert football player perceives more and more fine-grained possibilities for action and how they affect the unfolding of the situation. The skilled player perceives the gap between the two defenders as "for-running" in the context of a soccer game in which a teammate is advancing on the left flank. During a defensive corner, the same gap might be perceived as "for-countering" and not solicit any direct action, but might instead ready the agent to make a move.

The central problem of interest for a cognitive science studying skilled action is, I believe, that of context-sensitive selective openness to only the relevant action possibilities. Cognitive science needs to explain how selective sensitivity to relevant affordances is shaped by context and previous experience in a way that realizes grip on the situation. Next, I will continue to argue that active inference, understood in the proper way, is the right kind of framework for such a kind of cognitive science.

I have argued that what the agent is "modeling" in a concrete situation is not so much the causal structure of the environment, but rather the relevant action possibilities that bring the agent closer to a self-generated optimum. Brain dynamics self-organize as to enact an action-oriented relevance-centered perspective on the world. When responded to this action-oriented perspective leads to interactions with the world that in turn lead to a new perspective in which other aspects of the environment stand out as relevant and so forth. This is the circularity at the heart of both the free-energy principle and of skilled action. What the agent needs to be modeling then is not the relation between sensory stimulation and the causal structure of the environment *per se*, but rather the relation between sensory stimulation and its ways of living/flourishing in an ecological niche with a particular action-related structure. The generative model of the agent is thus shaped by previous experience resulting in more and more subtle refinements to the context-sensitive relevance of available affordances.

This interpretation of active inference has a number of distinct features. First, it conceptually blurs the distinction between epistemic and purposive actions. Tending towards an optimal grip includes both running in a gap between defenders as well as looking to whether an anticipated pass is coming or not. There is no clear demarcation between the two. Second, it puts both perception and action in the service of tending towards a (partly) self-generated optimum and provides conceptual grounds for explaining where this optimum comes from. Rather than appealing to the need for truthful representations or the need for homeostasis, I appeal, in the case of humans, to the normative character of the socio-cultural practices in which the agent participates. It takes a skilled and enculturated agent to be sensitive to the relevant affordances of playing football. Last, and perhaps most importantly, it provides an account of intentionality without presupposing goals or intentions as unexplained explainers. Instead it tries to understand intentionality in terms of the agent's history of interactions with the environment based on a concern to improve grip (Bruineberg & Rietveld, 2014 (Chapter 1)). What is relevant is not calculated based on our inferred representation of the outside world and a desire or an intention, but rather directly shows up in the way a skilled agent perceives the world.

One might point out here that replacing an appeal to internal goals by the tendency towards an optimal grip merely shifts the problem of purposive action. With the notions of "skill" and "practice" I might just presuppose the goal-directedness that internal goals typically account for. I think this shift is warranted for two reasons. First of all, it breaks the problem of purposive action up in two parts: the purposiveness of the practice and the ability of the individual to more or less adequately take part in that practice, this leads to a different explanandum. Second, and more importantly for this paper, active inference, at least for humans, *requires* intentional practices for the acquisition of priors. The reason for this is intimately related to the 'crooked scientist argument': my priors need to be of an optimal world, not the actual world. As Friston notes:

> One straightforward way to acquire priors - over state transitions - is to marinate an agent in the statistics of an optimal world, as illustrated in(Friston et al., 2009). One might ask where these worlds come from. The answer is that they are created by teachers,parents and conspecifics. In robotics and engineering, the equivalent learning requires the agent to be shown how to perform a task. (Friston, Samothrakis, & Montague, 2012, pp. 524-525)

In other words, the developing infant is engaging with specific practices carefully set up so as to teach the infant the relevant aspects of its environment. This process of "education of attention" (Gibson, 1979, p. 254) shapes the individual's selective openness to affordances in a way specific to its form of life. Unfortunately, the theme of learning of optimistic priors is currently underdeveloped in the active inference literature, but cultural learning and participating in 'regimes of shared attention' (Ramstead et al., 2016) seem to hold the key to acquiring the right expectations. At any rate, what will not be sufficient is for an individual to learn the statistics of its actual environment, since the actual environment misses the optimality that active inference requires.

First and foremost, I hope to have shown that the Helmholtzian, the Ashbyian and the enactive-affordance based account are each very different interpretations of active inference. Unlike the Helmholtzian and the Ashbyian framework, the Merleau-Pontian framework is able to frontload the relevance problem in active inference. This is not to say that active inference solves the relevance problem, it should rather be a central problem to those studying active inference. Furthermore, active inference tacitly assumes the agent to be endowed with optimistic priors. This promotes the idea of the developing active inference agent as an apprentice rather than a scientist.

3.4 Sense of agency and predictive-processing

In the previous section, I have introduced different accounts of agency under the free-energy principle. In this section, I will discuss the notion of the *sense of agency*. Phenomenologically, the *sense of agency* is understood as the feeling of being the cause of one's actions; the feeling that accompanies intentional and agentive voluntary actions. This feeling might be present when I take a step forward, but not when I am being pushed forward (Gallagher, 2000). My starting point, in this section, will be an early, and interesting, proposal by Hohwy (Hohwy, 2007) to map the sense of agency onto predictive processes. Hohwy provides a clear functional role for the self in agency and bodily movement:

> An individual needs to be able to generate and intimately track motor commands in accordance with her desires and beliefs about the world. There must be a distinction available between changes in her body and in the environment that are due to her own agency and those changes that are due to other factors in the environment or her sensorimotor system. (Hohwy, 2007, p. 2)

In other words, first, the agent needs to track whether an intention to act in the world actually has the desired result, and, second, distinguish between sensations following from its own movement and from other causes. The latter is explained by predicting the sensory consequences of a self-initiated movement and comparing them with actual sensory (reafferent) feedback. In expected situations, the error-signal that will be passed on in the model will be precise, which leads to attenuation of reafferent feedback thereby giving rise to what Hohwy calls a 'sense of mineness' of the movement. In a sense, we are 'at home' in the movement because we can precisely predict the sensory consequences of the movement. In contrast, we can't in the same way precisely predict the sensory consequences of other people's movements as well as our own. Hence the sensory consequences arising from the other's movements are not attenuated and so we don't experience the same feeling of mineness. According to Hohwy, the feeling of mineness *colors* our experiences in such a way as to enable us to perceive "one's body as a locus of mental causation", and to understand "where the mind ends and where the world begins" (Hohwy, 2007, p. 2). In other words, the feeling of mineness is necessary to make sense of ourselves as agents acting in a world that makes sense.

A similar explanation might be constructed for a sense of self in case of perception. Perceptual inference depends on the disambiguation of self-caused and other-caused sensory stimulation: when moving around, I need to as it were "subtract" the influence of my own movements from percepts to be able to infer the state of the environment. Perceptual mineness is experienced when we are able to predict what we perceive: when we are able to understand the changes in the persistent, external world. Susan Hurley (1998) (based on Gallistel, 1980) provides a contrast class in which a man with paralyzed eye muscles tries to look to the right. While the eye does not move (the pattern on the retina stays the same), for the man the world appears to move to the right. The anticipated change in sensory input creates an experience in which the world seems to rotate in the direction of the anticipated glance.

There is an interesting link here between "perceptual mineness" and knowledge of so-called sensorimotor contingencies (O'Regan & Noë, 2001). If I am able to anticipate how my percepts are to change if I moved in a particular direction, I will gain both a sense of "perceptual mineness" in which I am "at home" in the situation, but also a sense of "perceptual presence". Hohwy is thus completely right to state that: "as you gain the world you gain a sense of self" (Hohwy, 2007, p. 7).

What, I have argued, is distinctive of the enactive-affordance based account developed in the previous section, is that it provides a skill-based account of intentionality without presupposing internally represented goals and intentions. That is to say, for a skilled agent the relevant solicitations show up in perception. This is importantly different from accounts of the sense of agency that start from attenuation of reafferent feedback. Using such models, the account of the sense of agency starts with a precise counterfactual hypothesis ("I have a cup of coffee in my hand") and the temporary attenuation of actual sensory input ("my hand is resting on the keyboard"). This then triggers the body to change the world so as to make the sensory input fulfill the counterfactual prediction ("I have a cup of coffee in my hand"). A sense of agency arises when the sensory input changes in the way I anticipate. However, this approach presupposes the adequate generation of counterfactual predictions, which, in the PP framework take over the role of intentions. This is a commonly used strategy in the motor control literature:

"[w]ill or intentions are external input parameters similar to task parameters" (Latash, 1996, p. 302 quoted from Dotov & Chemero, 2014), but it is a problem for any theory that wishes to give an exhaustive and complete account of the workings of the brain and our minds. I take it that both predictive-coding and the free-energy principle have these ambitions.

A related distinctive feature of the enactive-affordance based account compared to the rationalist and cybernetic accounts is its emphasis on subjectivity. As Evan Thompson writes in *Mind in Life* (Thompson, 2007, p. 81): "[N]aturalism cannot explain matter, life and mind, as long as explanation means purging nature of subjectivity and then trying to reconstitute subjectivity out of nature thus purged." Making skilled intentionality basic to our account implies highlighting the perspective and the concerns of the individual. On the Helmholtzian account all purposiveness is reducible to tending towards a truthful representation of the structure environment (or left external to the theory). On the cybernetic account, purposiveness is accounted for only to the extent that it pertains to the control of homeostatic essential variables. On the ecological-enactive account there is no such unifying account of purpose. Although, on this account, agency is understood in terms of tending towards grip on the situation, what actually counts as the optimum that the agent tends towards in acting, is generated by the system itself and is a function of the agent's history of interactions with the environment, embodied in the agent's generative model.

3.4.1 The primacy of the 'I can'

In this section, I wish to highlight an aspect of the sense of self that is, arguably, more basic than the sense of agency from the last section. The phenomenon that I am after is quite simple: when I pick up a cup, I do not experience my fingers, but I experience the cup. Still, in the experience of the cup my body is not totally transparent to me. My body is not given to me as an object, but rather it is the subject of my experience. Similarly, when I perceive the solicitation of the coffee, I experience the coffee through my bodily capabilities (e.g., the ability to drink from a mug). This sense of bodily self works at the level of motor intentionality or skill, i.e., as intentional activities involving our bodily, situational understanding of space and spatial features. As Gallese and Sinigaglia (2010) note:

> [T]he bodily self has to be primarily and originally construed in terms of motor potentiality for actions, inasmuch the nature and the range of such potentiality define the nature and the range of pre-reflective bodily self-awareness. (Gallese & Sinigaglia, 2010, p. 753)

Their claim is that I pre-reflectively experience my body while grasping the cup, not as an arm, not as a hand, but as a bodily power for action. The horizon of action possibilities that the agent encounters (a field of relevant affordances), structured according to the demands of the situation and the agent's abilities, coincides with a coherent self as a bodily power for action. Importantly, for Merleau-Pontyians the relation between the horizon of possibilities and the coherent self is already intentional through and through. The primary sense of engaging with the world is in a bodily and skillful way, or as Merleau-Ponty, inspired by Husserl, famously states: "Consciousness is in the first place not a matter of 'I think that' but of 'I can' "(Merleau-Ponty, 1945/1962, p. 137).

So, how does this conception of the self as a bodily power for action relate to active inference and FEP? We have seen in the section on the main tenets of the free-energy principle that the starting point for the free-energy principle is biological self-organization, formalized in terms of the minimization of surprisal. As such, the reliance on action is not accidental, but constitutive for the being of the agent:

> I model myself as embodied in my environment and harvest sensory evidence for that model. If I am what I model, then confirmatory evidence will be available. If I am not, then I will experience things that are incompatible with my (hypothetical) existence. And, after a short period, will cease to exist in my present form (Friston, 2011, p. 117)

If we take this view seriously then the animal *needs* to expect itself to be a coherent agent acting in the world. Constitutive of self-organization, and basic to the free-energy principle, is an agent with the capacity to selectively interact with its environment to fulfill metabolic needs (Schrödinger, 1944). In order to be a free-energy minimizing agent, then, an agent needs (in a constitutive sense) to expect itself as having the capacity to selectively act on its environment to fulfill metabolic *needs*. This expectation is not available to consciousness as a belief or hypothesis, but is rather embedded in the structure of the agent's generative model. The consequence of this is, I believe, that I encounter myself in the first place not in introspection, but in the way the world shows up to me as relevant: in the solicitations I encounter.

If we assume that the generative model constitutes the agent's perspective on its environment, then the free-energy principle dictates that this perspective is structured in a particular way. The agent needs to be able to act on the world and it needs to be able to act in ways that improve the agent's relationship to its environment. If there is phenomenal component to active inference (this might depend on the kind of animal, but see Bruineberg et al. (2016 (Chapter 2)) on FEP and the mind-life continuity thesis), it will include something like "I can move to improve". The world-side of this phenomenological structure might consist of solicitations that stand out as relevant, pointing towards an improved grip on the environment. The animal-side might very well be captured by Gallese and Sinigaglia's notion of a coherent self as a bodily power for action.

I take it that this self-structure precedes any account of the self in terms of action-monitoring or comparing of intentions, for such accounts already presuppose the very ability to act. The free-energy principle requires an account of the self that captures its reliance on actions that are aimed to improve the condition of the organism in its environment.

What I hope to have shown is that the self in active inference is not accessible as an explicit belief or encountered as a thing, but shows up in the way the agent is drawn to improve its grip on the situation. This is required by FEP since only if the agent is a model of its econiche will the agent be able to maintain itself as the kind of being it is.

3.5 Conclusion

In this paper, I have investigated three perspectives on "predictivism": the Helmholtzian, the Ashbyian and the enactive affordance-based account on active inference. What is exciting about the paradigm of "predictivism" is its attempt to unify a plurality of cognitive concepts such as value and reward to a common currency: priors, prediction and precision. However, this by no means establishes the truth of the brain as analogous to a scientific hypothesis-tester. In particular, there is a tension between accounts that stress self-organization and metabolic needs and those that stress hypothesis-testing. I have argued that if the brain is thought of as a scientist, it needs to be a crooked scientist (contra the Helmholtzian interpretation). The Ashbyian account is better situated to account for bodily needs, because it starts from homeostasis, allows for both interoception and exteroception, and their integration. Yet, the Ashbyian account has its limits as well since not all of our actions can be grounded by metabolic needs. Some of our actions even squarely oppose our metabolic needs. I take it that theorists of active inference can draw important lessons from Merleau-Ponty's philosophy. Most notably from the kind of skilled action that Merleau-Ponty calls the tendency towards an optimal grip on a situation and the extent to which an animal brings forth its own world. I hope to have shown that even if one does not care about the merger of phenomenology and cognitive science, the Merleau-Pontian perspective on active inference still allows one to derive a number of research questions that are not easily derived from the other accounts presented. It is specifically able to frontload the question of relevance, and how an agent is able to select its own action possibilities given its previous history of interactions with the environment. Even purely behaviorally these are important questions that need to be highlighted in research on active inference.

It might seem odd to integrate neuroscience and phenomenology in the way attempted in this paper.[6] For one, the phenomenological tradition is often taken to be at odds with naturalism. For the moment, it suffices to understand this

---

[6] Thanks to an anonymous reviewer for pressing me on this point.

approach as taking inspiration from Merleau-Ponty, without claiming to actually be completely in line with his philosophy. A more radical thesis, to be defended in a future paper, is that if Merleau-Ponty were to be alive today, he would be a philosopher of complex systems theory.

Another point that requires some expansion is the connection between neuroscience and phenomenology. Much of contemporary phenomenology-friendly neuroscience takes a phenomenon of interest (such as the 'sense of self') and then tries to find the neural correlates of that phenomenon. The current paper stands in a rather different tradition: it attempts to develop a coherent and encompassing conceptual framework for skilled action including its neuroscientific, phenomenological and normative components. The comparison between the Helmholtzian, the cybernetic and the enactive affordance-based accounts layed out in this paper is not only about which account best fits the data, or providing knock-down arguments against one or the other, but about which one provides the most plausible, coherent and encompassing interpretation of active inference. The importance of such a framework is not just philosophical, but can have important practical ramifications. Consider one last time the case of standing under a too hot shower. To the modeler the option is always open to introduce an ad-hoc hyperprior that introduces an expectation that drives the agent away from the shower. The aim of a conceptual framework, like the Skilled Intentionality Framework, is to provide the right intuitions and theoretically justify choices made in modeling. We can understand moving away from the shower only if we think that active inference is about tending towards the most flourishing state of the animal-environment system, rather than the most likely causal structure of the environment per se.

The framework presented in this paper allows one to draw parallels between phenomenological structures and structures as they follow from theoretical biology. One of the great insights of Merleau-Ponty is that, as skilled humans, we have a prereflective bodily engagement with the world. Based on our concern of having grip on the environment, we are selectively sensitive to only particular solicitations that, when responded to lead towards grip on the situation. As the skilled agent perceives solicitations in the environment, it experiences itself as a bodily power for action: self and lived world evolve together. A similar structure follows from the free-energy principle: only by enacting its own viability conditions by expecting particular sensory information and acting to bring about the sensory information it expects, does the agent guarantee its own continued existence, and flourishing in its environment. The agent *needs* to model itself as an active agent with the capacity to selectively interact with its environment. This bodily self as power for action, this 'I can', has priority over any other account of the sense of agency, such as action-monitoring. For, unlike the others, this one does not presuppose intentions. It is in itself able to ground a specific kind of intentionality, which I have elsewhere labelled "Skilled Intentionality".

Chapter 4

General ecological information supports engagement with affordances for 'higher' cognition

Abstract

In this paper, we address the question of how an agent can guide its behavior with respect to aspects of the sociomaterial environment that are not sensorily present. A simple example is how an animal can relate to a food source while only sensing a pheromone, or how an agent can relate to beer, while only the refrigerator is directly sensorily present. Certain cases in which something is absent have been characterized by others as requiring 'higher' cognition. An example of this is how during the design process architects can let themselves be guided by the future behavior of visitors to an exhibit they are planning. The main question is what the sociomaterial environment and the skilled agent are like, such that they can relate to each other in these ways. We argue that this requires an account of the regularities in the environment. Introducing the notion of *general ecological information*, we will give an account of these regularities in terms of constraints, information and the form of life or ecological niche. In the first part of the paper, we will introduce the Skilled Intentionality Framework (SIF) as conceptualizing a special case of an animal's informational coupling with the environment. We will show how skilled agents can pick up on the regularities in the environment and let their behavior be guided by the practices in the form of life. This conceptual framework is important for radical embodied and enactive cognitive science, because it allows these increasingly influential paradigms to extend their reach to forms of 'higher' cognition such as long-term planning and imagination.

4.1 Introduction

The main theoretical concepts of ecological psychology are *affordance* and *ecological information* (Gibson, 1979). Affordances are possibilities for action provided by the environment. Ecological information is the set of structures and regularities in the environment that allow an animal to engage with affordances. An important open question is how far the applicability of the affordance-framework reaches. Important theoretical developments have been made concerning the concept of affordances in order for it to be applicable to more socio-cultural aspects of human and animal activities (Chemero, 2003, 2009; Heft, 2001; Reed, 1996; Rietveld & Kiverstein, 2014; Stoffregen, 2003). In this paper we investigate how we should conceptualize the corresponding notion of ecological information.

For some of Gibson's later followers (e.g., Turvey, Shaw, Reed, & Mace, 1981), information is to be understood in terms of lawful relationships between structures in the environment and patterns in light, vibrations, and the like. On this account, affordances are perceivable in virtue of the availability of structures in the ambient array (say, a light pattern) that lawfully specify the presence of affordances. At the same time, Gibson himself states that understanding the affordances of things and other humans "comprises the whole realm of social significance for human beings" (Gibson, 1979, p. 211). That is to say, ecological psychology is supposed to be able to deal with the full range of human social activities, not just with simple sensorimotor coordination.

There is a tension between this emphasis on lawful specification and the claim that affordances can be applied to the whole domain of human interactions. Some authors take a narrow approach, limiting affordances exclusively to those action possibilities that are lawfully specified by information (e.g., Turvey et al., 1981; Runeson & Frykholm, 1983; Golonka, 2015; Golonka & Wilson, 2016), while others understand affordances more broadly, relating them to the full range of our human form of life, structured by conventions, customs, socio-cultural practices, or other regularities (Costall, 1999; Heft, 2001; Rietveld & Kiverstein, 2014; Bruineberg et al., 2016 (Chapter 2); van Dijk & Rietveld, 2017). These two strands of research use different definitions of affordances (see Chemero, 2003; Rietveld & Kiverstein, 2014) and require a different account of information.[1] In this paper we focus on the human form of life in all its richness and therefore take the broader approach. We are concerned with the informational structures that are able to support the whole realm of skilled human activities.[2]

---

[1] This also means that the broader understanding of affordances can do justice to (situated) normativity (Rietveld, 2008a; Chemero, 2009) and the more restrictive (law-based) understanding cannot, which limits the extent in which the latter type of account is able to deal with many kinds of 'higher' cognition.

[2] It is an open and interesting question to what extent these informational structures for 'higher' cognition figure in the form of life of other animals as well. A comparative study is beyond the scope of this paper.

We[3] will argue that what we call 'the human form of life' is replete with (partially non-lawful) regularities that support our interactions with the environment. These regularities do not just support visually guided actions, but also activities that are traditionally seen as 'higher cognition' such as imagination and long-term planning.

In the first section of this paper, we present the Skilled Intentionality Framework (SIF) (Rietveld & Kiverstein, 2014; Bruineberg & Rietveld, 2014 (Chapter 1); van Dijk & Rietveld, 2017; Rietveld et al., in press), a philosophical, ecological-enactive approach to understand the situated and affective embodied mind. SIF follows the guidelines of radical embodied cognitive science (Chemero, 2009) to account for cognition, action, and perception in terms of dynamical systems theory and organism-environment coupling, and without invoking mental representations. In the second section of the paper we present different ways to analyse the structure of the human form of life and identify the notion of a constraint as a useful term in which to understand both lawful and non-lawful regularities. In the third section of the paper, we analyse how regularities in the form of life can support forms of cognition that are typically seen as forms of 'higher cognition'.

4.2 The Skilled Intentionality Framework

The basic phenomenon that we are interested in, in this paper, is how a skilled animal can coordinate its behavior with a complex and dynamical environment. By this we mean that the animal is informationally coupled to only a locally present aspect of its environment and still is able to coordinate its behavior with respect to distal aspects of the ecological niche. This requires an understanding of the relations between aspects of the environment such that a skilled agent can, based on its skills and learning history, coordinate adequately with it. Before we can move to the environmental regularities presupposed in this coupling, we first present the animal-environment coupling that forms the background of this paper. We will look at this organism-environment coupling in terms of three complementary perspectives: normativity, intentionality, and dynamics.

First, the animal's coupling with the regularities in the environment is about appropriate behavior: for example there is something distinctively inadequate about waiting for a train in the middle of a meadow. Thanks to an extended process of the education of attention, skilled actors have been introduced into their ecological niche by more experienced practitioners and are typically able to act in appropriate ways in concrete situations they encounter (Rietveld, 2008a). Second, we follow the phenomenological tradition of Heidegger and Merleau-Ponty in realizing that our skills give us a distinct mode of intentional access to the world. The skilled animal perceives its environment in terms of the action pos-

---

[3] In the following, the pronoun "we" will be used for any work authored by at least one of the authors of this paper: Jelle Bruineberg, Anthony Chemero and Erik Rietveld

sibilities that *matter* to it. Skilled access to relevant aspects of the world is the norm in cognition, not an exception (Noe, 2012; Ingold, 2011). Third, from a dynamicist perspective we can understand skilled agency in terms of the dynamics of animal-environment systems, including the neurodynamics embedded or "nested" in the dynamics of the entire brain-body-environment system (Chiel & Beer, 1997; Byrge, Sporns, & Smith, 2014).

The Skilled Intentionality Framework (SIF) attempts to integrate these three perspectives on an agent's interactions with the environment in one framework. Skilled intentionality is defined as the selective engagement with multiple affordances simultaneously. Some of the affordances we encounter in a specific situation go unnoticed, others solicit an action. Work by Rietveld (2008a, 2012b) and by Withagen et al. (2012, 2017) has shown that ecological psychology needs to distinguish between "affordances" and "invitations" or "solicitations".

We have defined affordances at the level of the form of life as a whole (the ecological niche), namely as relations between aspects of the sociomaterial environment and abilities available in form of life (Rietveld & Kiverstein, 2014). For understanding the engagement of a particular individual agent we need to look at the following two relata: the agent's ability and the aspect of the environment (Chemero, 2003, 2009). Affordances specify what is *possible* for agents to do. Solicitations, on the other hand, specify what stands out for the individual agent as *relevant* to do in a particular situation. While sitting at a desk writing a journal article, there are lots of things an academic could do: watering the plants, rearranging books, going home early. However, in the current context the solicitations are, if all goes well, limited to the keyboard, the screen, a cup of coffee, and a small pile of papers. These solicitations or attracting affordances involve an experienced tension of something that stands out to be done (Merleau-Ponty, 1945/1962; Dreyfus & Kelly, 2007). The theory of affordances therefore needs a theory of agency or skilled intentionality, of how an individual agent can be selectively open to only the relevant affordances in a given situation.

In Bruineberg and Rietveld (2014 (Chapter 1)), we have developed a largely Gibsonian and Merleau-Pontyian account of skilled intentionality. We have characterized skilled intentionality as the organism's tendency towards an optimal grip on a whole field of solicitations. In order to tend towards grip, we have to be *selectively open* to only the relevant affordances. We have argued that selective openness to affordances is constituted by the skilled animal's anticipatory dynamics, understood in terms of self-organizing states of action readiness, which are forms of action preparation (see also Bruineberg et al., 2016 (Chapter 2); Rietveld et al., in press; Frijda, 2007).

A key fact about skilled action is that of all the things an academic *could* do in her office, the vast majority of actions would be inappropriate, weird or forbidden. Her selective openness to affordances is appropriate with respect to a socio-cultural practice or *form of life* (Rietveld & Kiverstein, 2014; Wittgenstein, 1953, 1993). The form of life of a kind of animal consists of patterns in its behavior, i.e., relatively stable and regular ways of doing things. For humans, these regular patterns are manifest in normative behaviors and customs in communities. Human beings share not only biology, but also embedding in sociocultural practices. We share more or less stable ways of living with others (cf. '*feste Lebensformen*'; Wittgenstein, 1993, p. 397). The richness of affordances here and now for the academic are available because she partakes in the form of life of humans and that of academics, Macbook users etc. At the same time, her participation in these forms of life also limits which affordances can stand out as relevant in a given context. When she enters the university library, she adapts her gait, she will whisper rather than talk out loud and will switch her phone to silence-mode.

On the one hand, SIF attempts to provide an accurate conceptualization of skilled action; on the other hand, the question is how such episodes of skilled action are possible. Skilled intentionality, i.e., coordinating with multiple relevant affordances simultaneously, requires the animal to be sensitive to and adapt to regularities and structures that are there in the environment. In the rest of this paper, we explore how this selective sensitivity is possible, and how to conceptualize the regularities that are there in the environment. We will see that these regularities involve the whole form of life and comprise a range of different activities. This brings us, in the last section, a step closer to how skilled intentionality might be applicable to what is traditionally understood as 'higher cognition'.

4.3 The structure of the environment

The central question of this section is the following: what is the structure of the world at the ecological scale such that animals can couple with it in ways that result in organized, adaptive and creative behavior? An observation made by a long list of thinkers in philosophy, psychology and biology, is that the world at the ecological scale is quite unlike the world as typically studied by physics (Von Uexküll, 1934/1992; Merleau-Ponty, 2003; Jonas, 1966/2001; Gibson, 1979; Ingold, 2000, 2011; Thompson, 2007). One reason for this is the environment of the animal is not the equilibrium world of Newtonian physics and classical thermodynamics, but rather an environment full with optical, acoustic, vibrational and pheromonal fluxes and patterns, partly generated by the movement and the presence of the animal itself.

In *The Ecological Approach to Visual Perception* (1979), Gibson devotes nearly half the book on the structure of the environment such that animals can directly perceive it. For Gibson, the main concept involved in understanding the direct coupling of the organism with the affordances in the environment is that of *ecological information*. Ecological information is traditionally understood as a relation between energy in the medium (in the form of light, vibrations etc.), and the substances and surfaces of the environment. Structured light from, for instance, the sun bounces of the surfaces and structures in the environment so that at each point of observation, the light carries information about the structures it has bounced of off. Gibson goes to great pains to show that the information in the environment is rich enough to be able to specify the structures and surfaces in the environment and the perceiver's relationship to these structures. For Gibson:

> [t]he central question for the theory of affordances is not whether they exist and are real, but whether information is available in the ambient light for perceiving them (Gibson, 1979, p. 140).

For example, Sedgwick (1973) points out that the horizon cuts across objects at a height that is equal to the height of the point of observation. That is to say, whenever light is reflected to some point of observation from the horizon, and from some object between that point and the horizon, then the light at that point of observation can be used to perceive affordances related to the relative height of that object and the observer, such as the 'reachability' and the 'pass-under-ability' of an object for an observer.

Some of Gibson's most influential later followers (e.g., Turvey et al., 1981) focused on information provided by lawful relationships between structures in the environment and patterns in light, vibrations, and the like. However, importantly, lawful relations are not sufficient to account for the richness and diversity of the affordances available in the human form of life (Rietveld & Kiverstein, 2014). The overwhelming majority of affordances in human social relations are not lawfully specified by the energy in the environment, but are determined by conventions, customs, practices, or other regularities. For instance, a colleague might make coffee every morning and put it in a thermos. The thermos normally affords pouring coffee, despite the fact that occasionally the colleague is ill, and the thermos is empty, or contains the cold coffee of the day before. Although these regularities are not strictly law-like, since there can be exceptions (like in this case of illness), they do form the basis for the majority of our everyday skillful engagement with the environment in the human form of life. We will see that this is crucial.

As mentioned above, for Gibson, understanding the affordances of things and other humans "comprises the whole realm of social significance for human beings" (Gibson, 1979, p. 121). That is to say, ecological psychology is supposed to be able to deal with the full range of human social activities, not just with simple sensorimotor coordination. In the human form of life, these activities include creativity, long-term planning and imagination. Although the law-based notion of information is able to couple the light hitting our retina with the substances and surfaces in the environment, it is ill-equipped to couple us to the intricate patterns in human sociomaterial practices. Still, human living systems rely on these latter patterns or regularities for most of the distinctively human things that they do. For instance, the supermarket at the corner correlates reliably enough with the presence of bananas to enable banana-oriented behavior, even though every now and then the bananas happen to be out of stock there.

4.3.1 Lawful and general ecological information

In this paper, we will therefore distinguish between *lawful ecological information* and *general ecological information.* The former notion, especially as used by Turvey et al. (1981), pertains to a lawful regularity in the ecological niche between structure at a point p in a medium and an aspect of the environment at point q such that there is a 1:1 specifying relationship between p and q. When an animal is at p, it can be perceptually coupled to the affordances of q. We will call this restricted notion of ecological information *lawful ecological information.* In order to account for the sociomaterial character of the human ecological niche, we argue that a more general account of *ecological information* is required (henceforth: *general ecological information*).

General ecological information is any regularity in the ecological niche between aspects of the environment, x and y, such that the occurrence of aspect x makes the occurrence of aspect y likely. Because of the regular relation between the aspects of the environment x and y, general ecological information allows an animal to couple to a distal (i.e., not perceptually present in front of one's nose) aspect of the sociomaterial environment. General ecological information pertains to the ways in which aspects of the environment tend to occur together, like smoke and fire, an object and a shadow, or a pub and beer. Of course, this implies that the animal is able to couple to the relevant aspect of the environment. This depends on the abilities of an agent and the properties of the aspect of the environment i.e., a blind animal can't couple to a pattern in the light, and a human can only couple to the message on his phone by coupling to the light that shines off of it.

General ecological information is not limited to aspects of the environment that are perceptually present: something (say a bird of prey, aspect y) does not need to be perceptually present to get me ready to act on its affordances, because its shadow (the shadow of the bird, environmental aspect x) can reliably inform me about the presence of aspect y, even though in exceptional cases the shadow (aspect x) might be caused by a different aspect of the environment than aspect y (say for example aspect z, a kite). This example of the bird and its shadow also shows that the case of such *general ecological information* – due to the regularities in our ecological niche – is such that an aspect of the environment constrains (but does not necessarily specify lawfully) another aspect of the environment. *Lawful ecological information* also depends on regularities in our ecological niche and can best be seen as a special case[4] of *general ecological information*.

4.3.2 General ecological information and constraint

The notion of ecological information[5] (both lawful and general) pertains to regular relations between aspects of the environment (patterns, events, substances and surfaces etc.). The notion of ecological information is minimal, in the sense that it is present whenever there are regularities. It is informational in the sense that regularities that hold between aspects of the environment allow the animal *to be informed* about one aspect of the environment by the presence of another aspect. This sort of information does not imply meaning encoded in a signal. Information is nothing over and above these regularities between aspects of the environment.

Following situation semantics (Barwise & Perry, 1981, 1983) we will use the term "constraint" for the regularities between situations that reduce possibilities. Whenever there are constraints between types, there is information between tokens. An example that Chemero (2009) provides is the situation in which there is an unopened beer can on a table in a brightly lit room. Light from the source will reflect off the beer can and off the other surfaces in the environment. At each point in the room in which there is an uninterrupted path from the beer can, there will be light that has reflected of the beer can and is structured in a peculiar way. Due to the natural laws governing the reflection of light off surfaces and textures, the light at any such point in the room will be structured in a very particular way. In this case, there is a lawful constraint connecting the light-structure of type A to the beer-can presence of type B. In virtue of this constraint, the light structure at point p contains information about token beer-can-presence b (of type B) at some other point p'. Furthermore, and crucially for understanding our proposal, it is generally the case that unopened beer cans contain beer. Because of these conventional constraints governing cans and their contents, beer-can-presence b being of type B carries information about beer-presence c of type C. That is to say, because of the constraints and regularities, both physical and conventional, the light (ambient array) at some point in the room can carry information about the availability of beer.

---

[4] *Lawful ecological information* is a special case, or subclass, of *general ecological information* understood from the perspective of regularities: all regularities between aspects of the environment constitute *general ecological information*, but only regularities that determine (rather than constrain) the state of another aspect of the environment constitute *lawful ecological information*. As we will discuss below, this does not deny the fact that local lawful constraints enable the use of conventional constraints.

[5] Since this paper only deals with ecological information and not with other kinds of information, we will in the following use information and ecological information interchangeably.

Importantly, this sociomaterial regularity allows for exceptions: a mistake at the brewery might cause the can to be filled with water rather than beer. In the basic case, it is the *use of information* rather than *information* itself that is normatively evaluable.[6] In other words, there is nothing wrong (nor right!) about the light bouncing off a beer-can when it is accidentally filled with water. The light-structure of type A specifies beer-can presence no matter what: it carries information about beer-can presence without being able to be right or wrong. It is the *use* of information (by for example drinking the beer or stating that "this can contains beer") that is normatively evaluable. In the case of human made artefacts both the information itself and its use are normatively evaluable. If, by some other mistake in the factory, soda-cans end up with beer labels, the label *misrepresents* the contents of the can (even though the light bouncing of the can still specifies a beer-can). In summary, the use of information is always normative; artifacts that contain information are normative; energy arrays contain information but are not subject to norms.

Information pertains to the constraints that exist in the (sociomaterial) environment. These constraints can be necessary, such as the principle of non-contradiction, nomological or lawful, such as the laws of optics, or conventional, such as a thermos containing coffee or a supermarket selling bananas. While ecological psychology has typically focussed on finding lawful constraints, we are interested in the combination of lawful and conventional constrains that allow us to understand the relation between affordances, ecological information and the human form of life in all its complexity. The information induced by conventional constraints differs from the lawful constraint to the extent that it is not exception-free. The light structure of type fridge-presence does not infallibly specify the presence of beer in the fridge, but can still carry information about beer-presence if there is a constraint between the types of fridge-presence and beer-presence.

A related notion of constraint-based ecological information, also departing from Barwise and Perry (1983), is proposed by Sverker Runeson (1988, 1989). However, Runeson uses conventional constraints to argue that information is specific even though purely lawful constraints do not specify the layout of the environment.[7] In this paper, we are not committed to the claim that perception needs to be specific in order to guide behavior.

---

[6] Some authors (van Dijk, Withagen, & Bongers, 2015) state that "information about" implies normative evaluability (i.e., true or false, accurate or inaccurate). They therefore claim that conceiving of information independent of, and prior to, use is inappropriate. See **Section 4.3.4** for further discussion.

[7] One example Runeson discusses is the Ames-room (Wittreich, 1952), a spatially distorted room which gets "interpreted" as a rectangular room, giving rise to size illusions. Runeson argues that although alternative configurations are geometrically possible, they do violate conventional and pragmatic constraints such as that rooms have rectangular and horizontal floors and vertical walls.

Although constraints and regularities serve different purposes, they are not necessarily different things. To give an example: the convention of driving on the right puts constraints on the possible layout of intersections (e.g., the placing of road signs) and the location of the steering wheel in an automobile. The fact that this convention holds across continental Europe is a regularity that allows the exercise of the skill of driving, which was learned at one particular location, all over the continent.

Most of a skilled agent's interactions with the environment are far less explicit and articulable. Barwise and Perry (1983) provide the example that what makes someone a skilled basketball player is "her extensive *implicit* knowledge of the constraints that affect her, the ball, and the other things on the court" (p.98). Although, in order to clearly lay out our theory of information, we focus in this paper on relatively common sense and straightforward examples, such as beer cans and fridges, we recognize that ecological psychology is at its best when uncovering non-trivial regularities that agents use to coordinate their behavior, ranging from perceptual variables such as time-to-impact (Lee & Reddish, 1981) to multi-modal informational structures enacted by co-performers that constrain the behavior of interacting jazz musicians (Walton, Richardson, Langland-Hassan, & Chemero, 2015).

4.3.3 Constraints, information and form of life

Now let us try to apply the above account of information to the notion of a form of life. As we have seen before, the form of life of a kind of animal consists of patterns in its behavior, i.e., relatively stable and regular ways of doing things. The notion of a "form of life" allows us to capture the variety of socio-cultural practices within the human way of life. In the human form of life, the affordances available are related to the whole spectrum of abilities available in our human socio-cultural practices.

Wittgenstein shows the dependency of the human form of life on regularities as well as the interwovenness of the material and the socio-cultural aspects of our environment with an example of a familiar practice:

> [...] if things were quite different from what they actually are - if there were for instance no characteristic expressions of pain, of fear, of joy; if rule became exception and exception rule; or if both became phenomena of roughly equal frequency – this would make our normal language-games lose their point. – The procedure of putting a lump of cheese on a balance and fixing the price by the turn of the scale would lose its point if it frequently happened for such lumps suddenly grew or shrank for no obvious reason. (Wittgenstein, 1953, p. 48)

In our human environment, the conventional practices (say, of weighing cheese in order to price it), exist only in virtue of lawful stabilities in nature (most objects having a relatively stable weight over time).

The interwovenness of the material and the socio-cultural aspects of the form of life allows for a multiplicity of such dependencies. There is a regularity between the location of the steering wheel in a car, the material layout of roads and the socio-cultural norm of driving on the right. The fact that only driving left and right exist as stable norms might further be due to the fact that human and automobile locomotion takes place on surfaces and in gravitational circumstances that rule out passing on top or below. For scientific purposes, it might be interesting to investigate the historical or causal priority of some of these constraints; right-hand traffic existed before cars existed so the former constrained the latter and not visa versa. One might very well argue that the constancy of the gravitational force and the properties of cheese constrain and make possible the practice of weighing cheese.

But from the perspective of a participant in the practice, what exactly causes these regularities is irrelevant. All that is required is a sensitivity to the regularities that are there in the form of life such that, when for example approaching an intersection, the affordances that show up as relevant are in line with the regular ways in which things are done in the particular practice. To use another example of Barwise and Perry (1983), a skilled veterinarian does not need to possess a theory of how X-rays work in order to perceive that the dog's leg is broken. For the veterinarian it suffices to be sensitive to the constraint the state of the dog's leg puts on the pattern on the X-ray, even though this constraint itself is the result of a complex interplay between the dog's leg, electrons and the detector. In effect, given the physical, technological, and conventional practices in which the veterinarian engages, she can see through the X-ray to the dog's leg in the same way that she can see through the window to the trees outside. In both cases, there is information available to enable perception.

Many of our everyday activities are founded on conventional constraints. For example, at home one can look out of a window and see the roof of an arriving tram. Based on familiarity with the sociomaterial practice of tram 3 in Amsterdam, one can apprehend that this is the end point of the tram and that normally it will stop for a few minutes and everyone will get out, sometimes with the exception of the people working on the tram. There is the constraint between the arrival and the departure of the tram. There is a constraint between the clock of the tram driver and the fact that the tram is departing. There is a constraint between the clock of the tram driver and the clock hanging in the kitchen. As such, the sociomaterial environment that we inhabit is replete with constraints. It might be next to impossible (and unnecessary) to figure out how these constraints come about, but it is the existence of these constraints that enable us to coordinate our behavior with respect to aspects of the environment to which we are not sensorily coupled. Just as the veterinarian is able to perceive what needs to be done to the dog's leg while being sensorily coupled to the X-ray, we can perceive it is time to leave to take the tram by looking at a watch.

It is an open question how far back one can push the requirement for sensory coupling. Even if the tram were not yet actually present, we could still see the tram stop sign, we could still see the rails that lead it to, for example, the street called Ceintuurbaan in Amsterdam. In fact, nothing much is changed by the arrival of the tram because one can be familiar with this entire sociomaterial practice and attuned to the regularity of the arrival tram 3: another tram will arrive soon.

Even if the curtain is closed and we do not see the tram, then we can still see that we are in the familiar apartment, a place that is constantly placed next to the tram 3 stop. The clock in the kitchen might still inform us that we have to leave now to catch the next tram. Anyone with the right abilities and sensitivity to the regularities that allow one to reliably couple to the affordances will be able to coordinate with these distal aspects of the form of life in virtue of information about more local aspects, we suggest. So no direct sensorily coupling with an object (e.g., the tram) is necessary in order for an agent to be open to its affordances. Part of being skilled, of being at home in the situation, is exactly that of being able to adequately coordinate actions with respect to a great number of aspects of the environment in virtue of the presence of a small number of things that *reliably covary* with these aspects.

Needless to say, all of the constraints just mentioned are fallible, and none are lawful: the clock might be fast, the tram might be late, or the tram stop might be temporarily relocated due to construction work. Still, crucially, normally they provide the regularities that allow one to skillfully coordinate behavior in a form of life.

There are a few things to note here. First, this notion of regularity is not inconsistent with the Gibsonian account of lawful ecological information. Lawful ecological information, understood traditionally as a 1:1 specifying relationship between the structure of the light and the substances and surfaces of the environment, is a special case of the notion of general ecological information as we develop it here. When you grasp the cup of coffee in front of you, the structure of the light at some point p might fully specify the location and structure of the cup, such that it affords grasping. What we want to resist, however, is that affordances are limited *exclusively* to cases where we are lawfully coupled to an affordance.

For a skilled agent at home in her ecological niche, it is sufficient to be coupled to some relevant aspect of the sociomaterial practice even when that aspect is not present in the current environment. Even though we are not visually coupled to the beer inside the unopened beer can, the beer inside can still solicit us. Even with the curtains drawn, the tram can still afford catching due to the regularities in the form of life mentioned above.

Second, note that the account of regularities that we propose is an interesting case of niche construction: many of the constraints and regularities we have talked about in this section are products of human invention. The construction of clock time, especially the construction of time zones, led to a great explosion of things happening simultaneously or sequentially. A mobile phone can tell us when the tram arrives, even a tram in another city. It is an impressive technological feat that a pattern of light and sound at a point of observation somewhere else in a faraway country can appear in real time on a laptop in a Skype conversation. That is to say, it is not just the construction of material structures that change our ways of living, but also actively inducing patterns that make events happen at regular moments. Within our framework, the regularities of time keeping induce new constraints between situations (cf. van Dijk & Withagen, 2016). Without having to look out of the window, the light bouncing off the clock in the kitchen informs us that we should leave now in order to catch the tram and make it to the appointment on time.

Third, we want to be clear that the abilities that allow us to be informed by conventional constraints are not independent of lawful constraints. Indeed, the use of conventional constraints to couple to distal features of the world depends on the use of local lawful constraints. An apartment dweller can be coupled to the distant tram via the conventions connecting its arrival to the printed schedule, the way that the schedule constrains the actions of the conductor, and the conventions according to which she and the train conductor set their clocks. But she can only be coupled to the tram via these conventional constraints if she is also lawfully coupled to, for example, the clock in her kitchen via the lawful, optical constraints governing the light reflected off the hands of the clock's hour and minute hands.

In this section, we have introduced the notions of *lawful ecological information* and *general ecological information.* Lawful ecological information pertains to the currently most influential Gibsonian notion, especially as used and developed by Turvey et al. (1981). It is defined as a lawful regularity in the ecological niche between structure at a point p in a medium and an aspect of the environment at point q such that there is a 1:1 specifying relationship between p and q. *General ecological information* is defined as any regularity in the ecological niche between different aspects of the environment (x and y) such that the occurrence of x makes y likely. The regularities in the ecological niche can be captured by the notion of *constraint.* Constraints are relationships such that something being in a particular state constrains the state in which something else can be. For example, an aluminum can having a beer logo constrains the possible contents of the can, or a shadow having a particular shape constrains the object from which the shadow originates.

The nature of these constraints can be the object of scientific research, but for the skilled agent it suffices to be sensitive to these constraints. Constraints might be brought about by complex causal structures (such as X-rays), by ecological laws (such as optics), conventions (driving on the right) or habits (beer in the fridge). In the human form of life all these constraints are meshed together. In **Section 4.4**, we will discuss how our notion of general ecological information can be used to understand episodes of 'higher' cognition, but first we would like to deal with a potential objection to our account.

4.3.4 Information and use

One problem that emerges for a constraint-based notion of information, as proposed by us in this paper, has been articulated by Withagen and Van der Kamp (2010). They argue that the extension of ecological information to include non-specifying constraints poses, among others, the problem that information cannot specify the object of perception. If a structure in the array covaries with multiple aspects of the environment, then the array itself cannot individuate the object of perception. In other words: why does the light structure of type fridge-presence specify beer-presence rather than, say, milk-presence, even though both are reliably present?

Withagen and Van der Kamp sketch two possible solutions to this problem. The first is to deny that information has to specify the perceptual object and to maintain that in fact the array carries information about all these aspects of the environment. Some internal process (such as an intention) then further determines which of these aspects is perceived. Withagen and Van der Kamp dismiss this option for they claim that a theory of information should, ideally, explain the object of perception and postulating internal process necessary to individuate the perceptual object violates the principles of ecological psychology.

An alternative solution, they hold, is to define information relationally and to deny that the array carries information independent of use (see alsovan Dijk et al., 2015). Following developmental systems theory (Oyama, 2000, 1985/2000), they argue that information is not intrinsic to a structure (such as an optic array or a strand of DNA), but relative to a system for whom that structure makes a difference. As such, "the impact of sensory stimuli is a joint function of the stimuli and the sensing organism; the 'effective stimulus' is defined by the organism that is affected by it" (Oyama, 1985/2000, p. 38).

We want to argue that although our particular constraint-based notion of information and the usage-based notion of information are substantially different, their resulting treatment of perception and action are not. First of all, Withagen and Van der Kamp (2010) make clear that they do not wish to deny "the highly structured energy patterns in the ambient arrays that animals can use" (p.158, our italics). For reasons mentioned above, they just take issue with equating information with such patterns. Constraints are necessary, but not sufficient for information, according to Withagen and Van der Kamp. We take it that both the usage-based and our constraint-based approach to information agree that perception is the result of a system of constraints interacting with a perceptual system sensitive to only some of these constraints. Constraint-based theorists like us define one of the relata (the constraints) as information; usage-based theorists define the relation itself as information. Consequently, for the former information is cheap, ubiquitous and user-independent, while for the latter it is rare, special and user-relative.

We suggest then that as long as the different notions of information are kept apart, they are not incompatible. Information-as-constraint pertains to the regularities in the environment independent of use by a particular individual and *does not* specify either the object of perception or what is relevant. Information-as-use pertains to the relational significance of an environment for an agent and *does* specify the object of perception and what is relevant. On our view, information and affordances are present in the ecological niche independent of the use by a particular agent. However, the animal cannot attune to all of these affordances at the same time, it needs to be *selectively* open or sensitive to only the relevant ones. While Withagen and Van der Kamp are worried that any notion of information that does not specify the object of perception requires some sort of "internal enrichment" in order to arrive at the object of perception, we, instead, argue that it just takes a process of selective openness to arrive at only the relevant affordances, or solicitations.[8]

Despite these differences, the user-based and our constraint based approach agree, contrary to traditional ecological psychology, that perception and action should be understood as a function of the agent-environment system as a whole. To understand an animal's directedness to its environment it is not sufficient to merely focus on the constraints and regularities in the form of life, but also to focus on how an agent is able to selectively be sensitive to or be invited by some affordances and not others. We have discussed in earlier work how to conceive of selective openness to affordances within radical embodied cognitive science and the Skilled Intentionality Framework (Bruineberg & Rietveld, 2014 (Chapter 1)).

4.4 Information for engagement with affordances for 'higher' cognition

In earlier work, we have discussed certain kinds of affordances for 'higher' cognition (Rietveld & Kiverstein, 2014; Rietveld & Brouwers, 2017; van Dijk & Rietveld, 2017). For example the possibility of judging correctly that this paper is written in English, or possibilities for imagining a future building. What is the nature of constraints and information involved in engaging with such an affor-

---

[8] The problem for the user-based account is to explain how the agent is selectively sensitive to some constraints and not to others, or, in their own words, how an agent modulates the invitiational character, or tunes the coupling strength, of some affordances and not of others without presupposing some sort of internal agent.

dance? Based on the theory developed so far in this paper, this kind of affordance is to be understood in the same way as affordances for mundane activities like grasping a cup.

How do we couple with what is absent in the sense of not immediately present to one of the senses? Notice that already in the very definition of an affordance there is an interesting notion of absence. Affordances are *possibilities* for action, and those actions must occur in the future. You *could* get your telephone out of your pocket now. Even when we are ready to act on an affordance, we are prepared for something that we *could* do, but what we could do is not yet done, so in a sense something is not yet there. There is no light bouncing off the future. This implies that there is something necessarily anticipatory or future-oriented in the perception of an affordance (Turvey, 1992). This absence only increases when the affordances become more complex.

We want to use affordances to talk about 'higher order' cognition in a way that views cognition as an unmediated contact of the skilled person with the regularities in the environment. The regularities apply at many different spatial and temporal scales, including at the scale of sociocultural practices. The individual's skills (most of which are acquired via a process of education of attention in sociocultural practices) provide access to the regularities of the world; some skills are primarily sensorimotor, such as grasping a cup, others are typically characterized as more abstract skills (e.g., imagination, but also grasping your own coffee cup from those on the table in front of you). Part of being skilled is knowing how to attune to the relevant pieces of information; i.e., coordinate with the relevant aspects of the environment.

4.4.1 Imagination: general ecological information and anticipatory states of action readiness

In this section, we will discuss the regularities involved in a case of imagination. Let us consider the real-life example of an architect designing an art installation, The *End of Sitting*, for the Chicago Architecture Biennial, while having previous built a similar installation in Amsterdam.[9] The End of Sitting is a large architectural art installation in which visitors are invited to engage with a landscape of standing affordances. It is a rock-like structure that affords a variety of supported leaning, standing and hanging postures.[10] Moreover, due to the large variety of

---

[9] This section discusses one kind of imagination as used by architects in a design process. There might be other forms of imagination, such as for example daydreaming, that require a slightly different analysis.

[10] Many human activities can be performed skillfully exactly because they rely on carefully designed niches. However, the design process of a new aspect of a niche itself (for example, a landscape of affordances for supported standing) is neither a blind process, nor a process in which one of the architects has the end result in mind and merely realizes it in the world (Ingold, 2013). Rather, the construction of a new aspect of a niche is itself a result of skilled practices

positions offered and the temporary comfort offered by each of them, the installations affords switching positions about every 20 to 60 minutes. (Rietveld et al., 2015; Withagen & Caljouw, 2016; Rietveld, 2016).

In designing the installation for Chicago, the architects are able to anticipate how the installation will be used because they have perceived how it is used in Amsterdam, both by observing visitors and by experiencing it themselves. Their imagining was done in all sorts of ways (e.g., by making sketches, cardboard models, drawings, images, standing in live-size mock ups, etc.). Because both the people in Chicago and the people in Amsterdam have similar bodies and partake in the human form of life, the practices of leaning and standing in both places will be sufficiently similar. That is to say, there will be a constraint between the practice in Amsterdam and the practice in Chicago, which enables the architect to be informed about the practice in Chicago by observing the practice in Amsterdam. Even though the architects in Amsterdam are not in sensory contact with Chicago in any standard sense, they are able to anticipate how their installation will be used in virtue of regularities concerning human bodies and practices of standing that hold both in Amsterdam and Chicago (plus minor adjustments for differences in body posture).

The architects, located in Amsterdam, are in touch with the practice of standing and leaning in Chicago, in a way similar to the thirsty person, situated in a socio-cultural setting in which beer is a common beverage, who perceives a closed beer can is in touch with the beer. In both cases the agent is coupled to a local aspect of its environment: the studio and the fridge respectively, which, in virtue of constraints in the form of life bring the agent in touch with some distal aspect of the environment: the practice in Chicago and the beer in the fridge, respectively. This kind of imagination of the situation in Chicago is not qualitatively different from perceiving a fridge, or the cup of coffee in front of me. The differences between the beer can and the case of imagination concern only the two relata of affordances: there are differences in the skills engaged and differences in the regularities of the environment involved. Note that these two relata (aspects of regular patterns in the sociomaterial environment on the one hand, and abilities on the other) are precisely the two relata of an affordance (Rietveld & Kiverstein, 2014). Ecological optics is just a special case of the class of regularities that can couple a skilled agent to its environment.

Imagination, as presented here, is turns out to be a form of anticipation, made possible by skill and regularities of the environment. One consequence of this position is that, contrary to how ecological information is typically understood, general ecological information is not necessarily tied to a medium that grounds the informational relation. Whereas the informational relation between

---

enabled by the skills of the architects and the affordance provided by the architecture studio (including for example the possibilities to build cardboard models and to make 3D-drawings) (Rietveld & Brouwers, 2017; Rietveld, Rietveld, & Mackic, 2015).

a point of observation and a beer can is grounded by the properties of the local medium, the informational relation between practices of standing in Chicago and in Amsterdam is not. Still, in the case of architectural practice, the information relation between a scale model and a point of observation is grounded in the local medium. The information connecting the scale model and the installation in Chicago is a feature of the sociomaterial practices of architects. It is an open question whether all cases of imagination, such as for instance a science fiction writer writing about a completely imaginary universe, can be understood in this way.

In sum, the example of the architects in Amsterdam anticipating how something will be used in Chicago shows that imagination amounts to dealing with locally absent users (i.e., distally present users) primarily on the basis of regularities in human behavior. Imagination is a form of anticipation made possible thanks to the presence of regularities in our sociomaterial surroundings. Hence, there is no need to over-intellectualize imagination by understanding it in terms of mental representations. Our basic claim here is that things tend to happen in regular patterns that occur both locally and distally. Being attuned to these regularities allows the agent to let its behavior be guided by these patterns. Skills can be applied in new situations/instances appropriately because some of these regular patterns are the same across situations. Part of being skilled is knowing how to attune to the relevant pieces of information; i.e., connecting skills with the relevant aspects of the local and distal environment. When we are attuned to the regularities in the environment, we are able to be selectively open to those affordances that improve our grip on the situation, including affordances for activities that people have typically categorized as 'higher' cognition. What couples us selectively with the world is the action-readiness of the skilled animal to be sensitive to environmental constraints and act on a relevant affordance.

4.5 Conclusion

In this paper, we have presented an account of ecological information that is able to ground an animal's skilled interactions with absent aspects of the environment. The other main contribution of this paper is to distinguish between *lawful ecological information* and *general ecological information.* While ecological psychology has typically focused on the former, emphasising lawful informational coupling between agent and environment, our more general notion of ecological information has broader applicability, especially when it comes to sociomaterial aspects of the environment and forms of so-called 'higher' cognition. We have defined *general ecological information* as any regularity in the ecological niche between different aspects of the environment (x and y) such that the occurrence of aspect x makes the occurrence of aspect y likely. The nature of this regularity can be captured by different kinds of constraints. Constraints can, among others, be resulting from complex causal structures (such as X-rays), laws at the ecological scale (such as optics), conventions (like driving on the right), or habits (such as having beer in the fridge). The existence of these regularities and constraints allows for the possibility of having the skill to be sensitive to these constraints; that is, these constraints allow situations to inform agents about what can be done.

On our view, constraints make information available in our form of life, independently of any particular observer/agent. They pertain to the ways in which aspects of the environment tend to occur together, like an object and its shadow or an invitation and a party. General ecological information is understood as a relation between aspects of the environment (e.g., a tree and its shadow), but it is not relative to a particular observer or agent. This does not contradict the fact that agents can generate certain kinds of information, for instance by moving.

We have discussed the Skilled Intentionality Framework (SIF). Skilled intentionality is coordinating with multiple affordances simultaneously. Affordances are relations between aspects of the (socio)material environment and abilities available in a form of life. As such they are there also when no observer is locally present, but they are dependent on the existence of the form of life. Only when a kind of animal with the right leg-proportions and abilities arose, for example, did the log afford sitting on for that kind of animal. Solicitations, the relevant affordances in the concrete situation, are dependent on a particular individual's state of action-readiness. So, whereas information and affordances are available in the environment of the form of life, solicitations are there relative to the concrete individual animal in the concrete situation.

What we have highlighted is the continuity between simple cases of engaging with affordances and behaviors that are typically categorized as 'higher' cognition, such as architectural design that deals with something that is currently absent but to be built in the future. In all these cases, it is the skillful attunement to the regularities that are there in the environment (lawful, conventional or sociomaterial) that allow the agent to be responsive to affordances. The conception of affordances as relations between aspects of the sociomaterial environment and abilities available in a form of life allows us to circumvent artificial dichotomies between sensorimotor coordination and 'representation-hungry problems' (Clark & Toribio, 1994) or the intelligibility of the interface between 'contentless' basic minds and enculturated and linguistic forms of cognition (Hutto & Myin, 2013). Skilled agents are perfectly able to engage with and think about absent objects and sociomaterial practices, as long as they connect in regular ways with their environment.

However, one might wonder whether we have provided an ecological-enactive account of 'higher' cognition or have shown that what one might have thought of as 'higher' cognition is actually 'lower' cognition (i.e., only different in degree from animal cognition). Formally similar discussions occur in, for example, the field of artificial intelligence where a particular behavior (say, human-level competence in the game of Go) is a sign of genuine intelligent behavior until a computer can actually do it, after which it is understood as merely a case of pattern recognition.

We take it that 'dealing with the absent' and 'dealing with sociocultural norms' licenses the claim that in this paper we are 'dealing with 'higher' cognition'. We consistently use scare quotes for 'higher' cognition exactly because our work shows that there might not be a hard demarcation between 'lower' and 'higher' cognition. What matters for understanding both forms of cognition is that they are forms of skilled action or, more precisely, skilled intentionality. Dealing with the absent shows up already in the anticipatory character of engagement with affordances and can be extended to cases like imagination and architectural design, situated in socio-cultural practices. General ecological information does not only link the ambient array to structures or situations in the local environment, but also links between situations that are distant from one another in space and time.

Chapter 5

Free-energy minimization in joint agent-environment systems

Abstract

The free-energy principle is an attempt to explain the structure of the agent and its brain, starting from the fact that an agent exists (Friston & Stephan, 2007; Friston, 2010). More specifically, it can be regarded as a systematic attempt to understand the 'fit' between an embodied agent and its niche, where the quantity of free-energy is a measure for the 'misfit' or disattunement (Bruineberg & Rietveld, 2014 (Chapter 1)) between agent and environment. This paper offers a proof-of-principle simulation of niche construction under the free-energy principle. The key point of this paper is that the minimum of free-energy is not at a point in which the agent is maximally adapted to the statistics of a static environment, but can better be conceptualized as an attracting manifold within the joint agent-environment state-space as a whole, which the system tends toward through mutual interaction. We will provide a general introduction to active inference and the free-energy principle. Using Markov Decision Processes (MDPs), we then describe a canonical generative model and the ensuing update equations that minimize free-energy. Next, we apply these equations to simulations of foraging in an environment; in which an agent learns the most efficient path to a pre-specified location. In some of those simulations, unbeknownst to the agent, the environment changes as a function of the activity of the agent (i.e., unintentional niche construction occurs). We will show how, depending on the relative inertia of the environment and agent, the joint agent-environment system moves to different attracting sets of jointly minimized free-energy.

5.1 Introduction

What does it mean to say that an agent is adapted to - or ‘fits’ - its environment? Strictly speaking, in evolutionary biology, fitness pertains only to the reproductive success of a phenotype over evolutionary time-scales (Orr, 2009). However, this presupposes that an animal is “fit for purpose”; namely, to stay alive long enough to reproduce, given the statistical structure of its environment. On developmental time-scales, the animal comes to fit the environment by learning the statistics and dynamics of the ecological niche it inhabits. In other words, it acquires the skills to engage with the action possibilities available in its niche. On time-scales of perception and action, an organism improves its fit, or grip (Bruineberg & Rietveld, 2014 (Chapter 1)), by selectively being sensitive to the action possibilities, or affordances (Gibson, 1979; Rietveld & Kiverstein, 2014) that are offered by the environment.

Agents can not only come to fit their environments, but environments can come to fit an agent, or a species. For example, earth worms change the structure and chemical composition of the soil they inhabit and as a consequence, inhabit radically different environments in which they are exposed to different selection pressures - compared a previously uninhabited piece of soil (Darwin, 1881). In evolutionary biology, the process by which an agent alters its own environment to increase its survival chances is better known as “niche construction” (Odling-Smee, Laland, & Feldman, 2003). This leads to a feedback mechanism in evolution, whereby a modification of the environment by a member of a species alters the selection pressures working on the species’ members.

The aim of this paper is to discuss and model niche construction in the context of active inference and the free-energy principle (Friston & Stephan, 2007). The free-energy principle is a principled and formal attempt to describe the ‘fit’ between an embodied agent and its niche, and to explain how agents perceive, act, learn, develop and structure their environment in order to optimize their fitness, or minimize their free-energy (Friston & Stephan, 2007; Friston, Daunizeau, Kilner, & Kiebel, 2010). Whereas in evolutionary biology, this fit is primarily understood in terms of the optimization of heritable characteristics that enhance the agent’s probability of reproduction, the free-energy principle extends this fit to include multiple timescales, ranging from the optimization of neuronal and neuromuscular activity at the scale of milliseconds to the optimization of phenotypes over evolutionary timescales (Friston, 2011, Figure 10). The ‘fit’ between agent and environment is characterized by the information-theoretic quantity of (variational) free-energy.

Work on the free-energy principle has focused largely on epistemic perception and action (Friston, Adams, et al., 2012), goal-directed behavior (Pezzulo, Rigoli, & Friston, 2015), learning and exploration (Schwartenbeck, FitzGerald, Dolan, & Friston, 2013) and development (Friston, Levin, et al., 2015) (see Table 1 of Friston et al. (2017) for an overview). In this paper, we will apply the free-energy principle to *niche construction.* For the purposes of the current paper, we need to make a distinction between what one might call 'intentional' and 'unintentional' niche construction. In intentional niche construction, the agent alters its environment in order to obtain a desired goal or prior preference. For example, workers might set out to construct a sidewalk or a trail. In unintentional niche construction, environmental change is a byproduct of other sorts of behavior. For example, rushing on their way to work, people might cut the corner of the path through the park. While initially this might almost leave no trace, over time a path emerges, in turn attracting more agents to take the shortcut and underwrite the path's existence. Such 'desire paths'[1] are fascinating examples of unintentional niche construction and their emergence is a key focus of this paper.

The 'fit' between the agent and its environment can be improved both by the agent coming to learn the structure of the environment and by the environment changing its structure to better fit the agent. This gives rise to a continuous feedback loop, in which what the agent does changes the environment, which changes what the agent perceives, which changes the expectations of the agent, which in turn changes what the agent does (to change the environment). The deeper (philosophical) point here is that the minimum of free-energy is not (necessarily) at a point where the agent is maximally adapted to the statistics of a given environment, but can better be conceptualized as a stable point or, more generally, an attracting set of the *joint agent-environment system.*

The attracting set on which an agent-environment system settles will depend upon on the malleability of both the agent and the environment. In the limiting case of a malleable agent and a rigid environment, this amounts to learning. In the other limiting case of a rigid agent and a compliant environment, we find niche construction (making the world conform to one's expectations). In intermediate cases, both the agent and the environment are (somewhat) malleable. Importantly, as we will see later on in this paper, the malleability of the agent and the environment can be given a concise mathematical description; in terms of the prior beliefs encoded by concentration parameters of a Dirichlet distribution. These prior beliefs reflect the influence sensory evidence has on learning. In other words, they determine the 'learning rate' or 'inertia' of both the agent and the environment. These learning rates[2] embody the evolutionary and developmental history of an agent (the stability of the niche an agent evolved in) and the type of environment involved.

In brief, the active inference formulation described below offers a symmetrical view of exchanges between agent and environment. The effect of the agent on the environment can be understood as the environment 'learning' about the agent

---

[1] The Dutch term "olifantenpad" ("elephants' path") characterizes the nature of these paths in an imaginative way.

[2] One might be inclined to associate the agent with a learning rate and the environment with 'mere' inertia. Formally, however, we treat the agent and the environment equivalently, both parameterized by concentration parameters.

through the accumulation of ecological legacies (Laland, Matthews, & Feldman, 2016). This perspective is afforded by the basic structure of active inference that rests upon the coupling between a *generative process* (i.e., environment) and a *generative model* of that process (i.e., agent). The emergent isomorphism between the process and model means that there is a common phenotypic space that is shared by the environment and agent. On this view, the environment acts upon the agent by supplying sensory signals and senses the agent through the agent's action. Mathematically, the environment accumulates evidence about the generative models of the agents to which it plays host. This symmetry plays out in a particular form, when we consider the confidence or precision placed in the prior beliefs of the environment and agent - and the effect the relative precisions have on the convergence or (generalized) synchronization that emerges as the agent and environment "get to know each other".

In what follows, we will provide a general introduction to active inference and the free-energy principle. Using Markov Decision Processes (MDPs), we then describe a canonical generative model and the ensuing update equations that minimize free-energy. We then apply these equations to simulations of foraging in an environment; in which an agent learns the most efficient path to a pre-specified location. In some of those simulations, unbeknownst to the agent, the environment changes as a function of the activity of the agent (i.e., unintentional niche construction occurs). We will show how, depending on the relative inertia of the environment and agent, the joint agent-environment system moves to different attracting sets of jointly minimized free-energy.

5.2 The free-energy principle and active inference

The motivation for the free-energy principle is to provide a framework in which to treat self-organizing systems and their interactions with the environment. Below, we will briefly rehearse the arguments that lead from the desideratum of self-organization to the minimization of free-energy: for details, see Friston and Stephan (2007), Friston (2011) and, in more conceptual form Bruineberg et al. (2016 (Chapter 2)).

The starting point of the free-energy principle is the observation that living systems maintain their organization in precarious conditions. By precarious we mean that there are states an organism *could* occupy but at which the organism would lose its organization. Hence, if we consider a state space of all the situations an organism can be in (both viable and lethal) we will observe (by necessity) that there is a very low probability of finding an agent in the lethal parts of the state space and a high probability it occupies viable parts. Although which states are viable is dependent on the kind of animal one observes; namely, on their *characteristic states*.

We assume the agent has sensory states that register observations or outcomes $\tilde{o}$, where outcomes are a function of the state of the agent's environment, or hidden states, $\tilde{\boldsymbol{s}}$. These states are called "hidden" because they are "shielded off" from internal states by observation states. For an adaptive agent, its sensory states support a probability distribution $P(\tilde{o})$ with high probability of being in some observation states, and low probability of being in others, where - in analogy with the hidden state - frequently occurring outcome states are associated with viable, characteristic states and very rare outcome states are associated with potentially lethal states (see **Table 5.1** for notation, we will denote actual states in the environment with bold face $\tilde{\boldsymbol{s}}$, and states the agent expects in the environment using normal script $\tilde{s}$). Given the distribution $P(\tilde{o})$, one can calculate the surprisal (unexpectedness) of a particular observation o: $-\ln P(o)$. Observations that are encountered often, or for a long time, will have low surprisal, while outcomes that are (almost) never observed will have very high surprisal.

One expects a certain degree of recurrence in the states one finds any creature in. Take, for example, a rabbit: the typical situations a rabbit finds itself in might be eating, sheltering, sleeping, mating etc. It will repeatedly encounter these states multiple times throughout its life. Under mild (ergodicity) assumptions, the frequency with which we expect to find the rabbit in a particular state over time is equal to the probability of finding the rabbit in that particular state at *any* point in time. Ergodicity implies that the average surprisal over time is equal to the expected surprisal at any point in time, or mathematically:[3]

$$\sum_{s} -p(\boldsymbol{s}) \ln p(\boldsymbol{s}) = \sum_{t}^{T} -\frac{1}{T} \ln p(\boldsymbol{s_t})$$

5.2.1 Free-energy and self-organization

So far, we have adopted a descriptive point of view, starting from an adaptive agent. We can now turn from the descriptive statement - that adaptive agents occupy a restricted (characteristic) part of the state space with high probability - to the normative statement that in order to *be* adaptive, it is sufficient for the agent to occupy a characteristic part of the state space, which (by definition) must be compatible with the characteristic states of the agent in question. For example, the human body performs best at a core body temperature around 37°C. When measuring the temperature of a human, one expects to measure a core body temperature around 37°C, while measuring a body temperature of 29°C or 41°C would be very surprising and indicative of a threat to the viability of the

---

[3] Throughout this paper we will assume discrete time steps and categorical (discrete) states and outcomes.

agent. For adaptive temperature regulation then, it is sufficient to minimize the surprisal of observational states $\tilde{o}$ with respect to a probability distribution $P(\tilde{o})$ peaking at those temperature values that are characteristic of human bodies.

The observational states $\tilde{o}$ and the probability distribution $P(\tilde{o})$ serve to make the surprisal of an observation $-\ln P(\tilde{o})$ accessible to the agent. The ecologically relevant question for the agent is however how to minimize the surprisal of observations. Minimization of surprisal can only be achieved through action, be it by acting on the world (for example by moving into the shade) or changing the body (for example by activating sweat glands). That is to say, the agent needs to predict how actions $\boldsymbol{u}$ impact on observational states o. More often than not, the impact of control or active states $\boldsymbol{u}$ will be mediated by the hidden state of the environment $\boldsymbol{s}$: the action that reduces surprisal of temperature sensors depends on where the agent can find shade. Moreover, in many cases, surprising observational states can only be avoided by eluding particular hidden states in the environment preemptively. For example, a mouse can avoid being eaten by a bird of prey (a highly surprising state of affairs for a living mouse), by avoiding hidden states in which a bird of prey can see it. In turn, the diving bird causes a particular observation (a fleeting shadow, i.e., a sudden decrease in light intensity on its sensory receptors). The mouse therefore *needs* to treat the observation generated by a bird of prey as an unlikely state and avoid it by acting. Whether a particular observation is surprising or not therefore depends on the hidden state of the world that might have caused the observation. Crucially, in order to minimize the surprisal of observations, the agent also needs to be able to predict the consequences of its actions on the environment.

The surprisal of observations is therefore the marginal distribution of the joint probability of observations, marginalized over hidden states and policies the agent pursues:

$$-\ln P(\tilde{o}) = -\ln \sum\nolimits_{s,u,\theta} P(\tilde{o}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{u}}, \boldsymbol{\theta})$$

The probability distribution $P(\tilde{o}, \tilde{\boldsymbol{s}}, \boldsymbol{u}, \boldsymbol{\theta})$ is known as the *generative process* (where $\boldsymbol{\theta}$ represents a set of parameters), denoting the actual causal, or correlational, structure between action states $\tilde{\boldsymbol{u}}$, hidden states $\tilde{\boldsymbol{s}}$, and observation states $\tilde{o}$, parametrized by $\boldsymbol{\theta}$. Importantly, the agent only has access to a series of observations $\tilde{o}$ and not to hidden states $\tilde{\boldsymbol{s}}$ and actions $\tilde{\boldsymbol{u}}$. This means it cannot perform the marginalization above; instead we assume the agent uses a *generative model* $P(\tilde{o}, \tilde{s}, \pi, \theta)$, denoting the agent's expectations about the causal structure of the environment (generative process) and the policies it pursues.

Expression		Description
$P(\tilde{o}, \tilde{s}, \pi, \theta)$		*Generative model (agent):* joint probability of observations $\tilde{o}$, hidden states $\tilde{s}$, policies π, and parameters θ. Returns a sequence of actions $\boldsymbol{u_t} = \pi(t)$
$\boldsymbol{P}(\tilde{o}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{u}}, \boldsymbol{\theta})$		*Generative process (environment):* joint probability of observations $\tilde{o}$, hidden states $\tilde{\boldsymbol{s}}$, actions $\boldsymbol{u}$, and parameters $\boldsymbol{\theta}$. Generates observations: $o_t = \boldsymbol{A s_t}$
θ	$= (A, B, C, D)$	Parameters of the generative model
$\boldsymbol{\theta}$	$= (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$	Parameters of the generative process
o_τ $\hat{o}_\tau$	$\in \{0, 1\}$ $\in [0, 1]$	Outcomes and their posterior expectations
$\tilde{o}$	$= (o_1, \ldots, o_t)$	Sequences of outcomes until the current time point.
s_τ $\hat{s}^\pi_\tau$	$\in \{0, 1\}$ $\in [0, 1]$	Inferred hidden states and their posterior expectations, conditioned on each policy.
$\tilde{s}$	$= (s, \ldots, s_T)$	Sequences of inferred hidden states until the end of the current trial.
$\boldsymbol{s_\tau}$	$\in \{0, 1\}$	Actual hidden state, (analogous notation for posterior and sequences).
π $\hat{\pi}$	$= (\pi_1, \ldots, \pi_k) : \pi \in \{0, 1\}$ $= (\hat{\pi}_1, \ldots, \hat{\pi}_k) : \hat{\pi} \in [0, 1]$	Policies specifying action sequences and their posterior expectations.
$\boldsymbol{u_t}$	$= \pi(t)$	Action or control variables
A $\bar{A}$	$\in [0, 1]$ $= \psi(\alpha) - \psi(\alpha_0)$	Likelihood matrix mapping from inferred hidden states to outcomes and its expected logarithm.
$\boldsymbol{A}$ $\bar{\boldsymbol{A}}$	$\in [0, 1]$ $= \psi(\boldsymbol{\alpha}) - \psi(\boldsymbol{\alpha_0})$	Likelihood matrix mapping from actual hidden states to outcomes and its expected logarithm.

B^{π}_{τ}	$= p(s_{i,t+1} \mid s_{j,t}, \pi)$	Transition probability for hidden states under each action prescribed by a policy at a particular time and its logarithm.
$\bar{B}^{\pi}_{\tau}$	$= \ln B^{\pi}_{\tau}$	
D	$\in [0,1]$	Prior expectation of the hidden state at the beginning of each trial.
U_{τ}	$= \ln p(o_{\tau}) \leftrightarrow p(o_{\tau}) = \sigma(U_{\tau})$	Logarithm of prior preference over outcomes or utility.
F_{π}	$= F(\pi) = \sum_{\tau} F(\pi, \tau) \in \mathbb{R}$	Variational free-energy for each policy.
G_{π}	$= G(\pi) = \sum_{\tau} G(\pi, \tau) \in \mathbb{R}$	Expected free-energy for each policy.
$\hat{s}_{\tau}$	$= \sum\limits_{\pi} \pi \cdot \hat{s}^{\pi}_{\tau}$	Bayesian model average of hidden states over policies
H	$= -diag(\breve{A} \cdot \hat{A})$	Vector encoding the entropy or ambiguity over outcomes for each hidden state.
$\hat{A}$	$= E_Q[\ln A] = \psi(\alpha) - \psi(\alpha_0)$	Expected outcome probabilities for each hidden state and their expected logarithms.
$\breve{A}$	$= E_Q[A_{ij}] = a \times \alpha_0^{-1}$	
$\alpha_{0,ij}$	$= \sum\limits_{i} \alpha_{ij}$	
$\psi(\alpha)$	$= \partial_{\alpha} \ln \Gamma(\alpha)$	Digamma function or derivative of the log gamma function.
W	$= \frac{1}{a_0} - \frac{1}{a}$	A matrix encoding uncertainty about parameters for each combination of outcomes and hidden states.

Table 5.1: Glossary of expressions

We can pause for a moment and discuss the implications of this separation between the *generative process* and the *generative model.* The generative process pertains to the actual structure of the world that generates observations for the agent. In contrast, the generative model pertains to how the agent expects the observations to be generated. The agent will intervene in the world under the assumption that its generative model is close to the generative process. The generative model and process meet at two places: the environment is causing the observation states of the agent, and actions are sampled from a distribution over policies, selected by the agent under its generative model (see **Figure 5.1**).

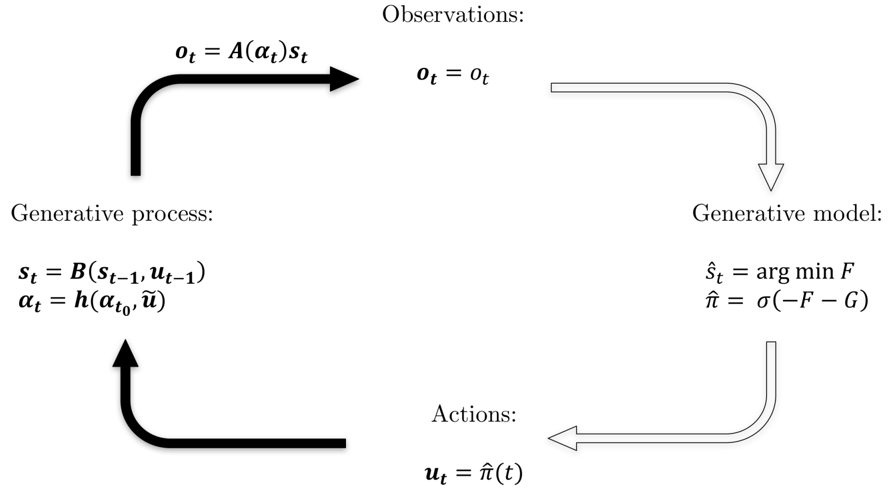


Figure 5.1: The generative process and model and their points of contact: the generative process pertains to the causal structure of the world that generates observations for the agent, while the generative model pertains to how the agent expects the observations to be generated. A hidden state in the environment $\boldsymbol{s_t}$ delivers a particular observation $\boldsymbol{o_t}$ to the agent. The agent then infers the most likely state of the environment (by minimizing variational free-energy) and uses its posterior expectations about hidden states to form a posterior over policies. These policies specify actions that change the state (and parameters) of the environment.

Note that, from the perspective of the agent, the agent uses its generative model to evaluate the surprisal of observations:

$$-\ln P(\tilde{o}) = -\ln \sum_{s,\pi,\theta} P(\tilde{o}, \tilde{s}, \pi, \theta)$$

However, although the agent has access to all the variables in the above equation, this marginalization is analytically intractable; so the minimization of surprise is not possible directly. Instead, one can consider an upper bound on surprise that can be evaluated and subsequently minimized; thereby explaining surprisal minimizing exchange with the environment in a way that can be plausibly instantiated in a living creature.

One can construct this upper bound by adding an arbitrary distribution $Q(\tilde{s}, \pi, \theta)$ to the surprisal term and using the definition of the expectation or expected value $E_{q(x)}[x] = \sum_x q(x) \cdot x$:

$$-\ln P(\tilde{o}) = \ln \sum_{s,\pi,\theta} Q(\tilde{s}, \pi, \theta) \frac{P(\tilde{o}, \tilde{s}, \pi, \theta)}{Q(\tilde{s}, \pi, \theta)} = -\ln E_{Q(\tilde{s},\pi,\theta)} \left[\frac{P(\tilde{o}, \tilde{s}, \pi, \theta)}{Q(\tilde{s}, \pi, \theta)} \right]$$

Using Jensen's inequality (following from the concavity of the log function), we then have the following inequality:

$$-\ln P(\tilde{o}) = -\ln E_{Q(\tilde{s},\pi,\theta)} \left[\frac{P(\tilde{o}, \tilde{s}, \pi, \theta)}{Q(\tilde{s}, \pi, \theta)} \right] \leq -E_{Q(\tilde{s},\pi,\theta)} \left[\ln \left(\frac{P(\tilde{o}, \tilde{s}, \pi, \theta)}{Q(\tilde{s}, \pi, \theta)} \right) \right] = F$$

The term on the right-hand side of the equation - the free-energy F - is therefore an upper bound on the term on the left-hand side of the equation, the surprisal of observations. In short, minimizing free-energy implicitly minimizes surprisal.

5.2.2 Free-energy and variational inference

The question then is how the minimization of free-energy can be achieved, and what this optimization entails. We have defined free-energy in terms of a generative model $P(\tilde{o}, \tilde{s}, \pi, \theta)$ and an arbitrary variational distribution $Q(\tilde{s}, \pi, \theta)$. The free-energy can be written in several forms to show what its minimization entails, specifically:

$$F(\tilde{s}, \pi, \theta) = \underbrace{D_{KL}[Q(\tilde{s}, \pi, \theta)||P(\tilde{s}, \pi, \theta|\tilde{o})]}_{\textit{divergence}} - \underbrace{\ln P(\tilde{o})}_{\textit{log evidence}}$$

This formulation shows the dependency of the free-energy on beliefs about the hidden states implicit in the variational distribution. Since the log evidence does not depend on $Q(\tilde{s}, \pi, \theta)$, optimizing the variational distribution to minimize free-energy means that the divergence from the posterior $p(\tilde{s}, \pi, \theta|\tilde{o})$ is minimized. This makes $Q(\tilde{s}, \pi, \theta)$ an approximate posterior, i.e., the closest approximation of the true posterior $P(\tilde{s}, \pi, \theta|\tilde{o})$. This highlights the relationship between free-energy minimization and theories of perception as Bayesian inference (Gregory, 1980). Furthermore, since the KL-divergence is always greater than zero, minimizing free-energy makes it a tight upper bound on the negative log evidence or surprisal.

Whether the exact minimization of free-energy is feasible depends on the generative process and generative model. Typically, simplifying assumptions need to be made about the form of the variational distribution, resulting in approximate rather than exact inference. The most ubiquitous assumption about the variational distribution is that it can be factorized into marginals. This is known as the mean field approximation (Opper & Saad, 2001). The only parameters θ that will vary in this paper are the parameters of an observation matrix $A \subset \theta$ and we can deal with a variational distribution of the form:

$$Q(\tilde{s}, \pi, A) = Q(\pi)Q(A)\prod_{t}^{T} Q(s_t|\pi)$$

The challenge now is to find the approximate posterior $\tilde{Q}$ that minimizes free-energy given a series of observations $\tilde{o}$ and the generative model $P(\tilde{o}, \tilde{s}, \pi, \theta)$. In other words, we want to find those $\tilde{Q}$ such that:

$$\begin{aligned} Q(\tilde{s}, \pi, A) &= \arg\min_{Q} F \\ &\approx p(\tilde{s}, \pi, A|\tilde{o}) \end{aligned}$$

This will provide update equations that formalize the exchange between the agent and its environment that is consistent with its existence, through a variational process of self-organisation. Due to the way the variational distribution is factorized, each factor can be optimized separately. The specific update equations specified in the next section are obtained by taking the functional derivative of the free-energy with respect to each factor and solving for zero. We can then construe a differential equation whose fixed point coincides with this solution, i.e., the minimum of free-energy. The result is a set of self-consistent update equations that converge upon the minimum of free-energy (see **Appendix B** and Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2016). Although not relevant for the current treatment, these equations have a lot of biological plausibility in terms of neuronal processes - and indeed non-neuronal processes involving cellular interactions: for further discussion, see Friston et al. (2017); Friston, Levin, et al. (2015). In short, if these variational constructs are the only way to solve a problem that is necessary to exist in a changing world, we can plausibly assume that evolution uses these constructs: more precisely, evolution it is itself a form of variational free-energy minimization.

5.2.3 Adaptive action and expected free-energy

Policies, or sequences of actions, do not alter the current observations, but only observations in the future. This suggests that the dynamics we are trying to characterize must be based upon generative models of the future. Furthermore, this means that an agent selects the policies that it expects will minimize free-energy in the future. This requires us to define an additional quantity, *expected* free-energy G, to ensure the agent acts so as to minimize the expected surprisal under a particular policy (i.e., pursue uncertainty-resolving, information-seeking policies that exploit epistemic affordances in their econiche). Above, we have defined the free-energy as:

$$F = E_{Q(\tilde{s},\pi,\theta)}[\ln Q(\tilde{s},\pi,\theta) - \ln P(\tilde{o},\tilde{s},\pi,\theta)]$$

In analogy with the variational free-energy, we can now define an expected free-energy under a particular policy π:

$$G(\pi) = \sum_t G(\pi,\tau)$$

$$G(\pi,\tau) = E_{\tilde{Q}}[\ln Q(s_\tau|\pi) - \ln P(s_\tau, o_\tau|\tilde{o},\pi)]$$

where $\tilde{Q} = Q(o_\tau, s_\tau|\pi) = P(o_\tau|s_\tau)Q(s_\tau|\pi)$. In other words: the expectation is taken under a counterfactual distribution $\tilde{Q}$ over hidden states and yet to be observed outcomes (and not over hidden states and policies, as was the case for the variational free-energy). Rearranging this expected free-energy gives (see **Appendix A**):

$$G(\pi,\tau) = D_{KL}[Q(o_\tau|\pi)||P(o_\tau)] + E_{Q(s_\tau|\pi)}H[P(o_\tau|s_\tau)]$$

Here, the second term is called ambiguity and reflects the expected uncertainty about outcomes, conditioned upon hidden states. The first term is the divergence between prior (i.e., preferred or characteristic) outcomes and the outcomes expected under a particular policy. This *Bayesian risk* or expected cost is the smallest for a policy that brings about observations that are closest to preferred observations. We can operationalise this sort of policy selection with a prior over policies that can be expressed as a Gibbs or softmax function of expected free-energy:

$$P(\pi) = \sigma(-G(\pi))$$

In short, the agent selects policies that it expects will minimize the free-energy of future observations. This is equivalent to minimizing Bayesian risk and resolving ambiguity.

So what does the minimization of free-energy entail in different contexts? In the limiting case of *perceptual inference* (where the agent cannot change the sensory array it is exposed to), free-energy is minimized by finding the hidden states $\tilde{s}$ that most likely generated observed sensory states $\tilde{o}$, under the agent's generative model of how they co-occur. This makes the recognition distribution $Q(\tilde{s})$ an approximate conditional distribution $P(\tilde{s}|\tilde{o})$. Here, the expected hidden states are the parameters of the variational distribution, which are generally considered to be internal states of the agent (e.g., neuronal activity).

When actions are allowed, but the agent has no preferences for particular states (*active inference without preferences*), free-energy is minimized by finding the hidden states $\tilde{s}$ that most likely generated observed sensory states $\tilde{o}$ and those actions are selected that minimize the ambiguity of observations given hidden states $P(o_t|s_t)$. This puts both action and perception in the fame of hypothesis-testing, or optimizing the Bayesian model evidence of an agent's model of its environment, licensing a Helmholtzian interpretation of the activity of the brain (Friston, Adams, et al., 2012).

However, when the agent is equipped with preferred sensory observations (*active inference with preferences*), the picture changes profoundly (Bruineberg et al., 2016 (Chapter 2)). Besides finding the hidden states $\tilde{s}$ that most likely generated observed sensory states $\tilde{o}$ and selecting those actions that bring about preferred outcomes; enabling it to elude surprising states of affairs. To give an intuitive example, the agent's current sensations might best be explained by the conjecture that he is standing under a shower that is too hot - a fairly unambigious signal. But, if all is well, standing under un uncomfortably hot shower is itself a highly surprising event. He will therefore reach for the tap to reduce the temperature and seek sensory evidence from the world that he is standing under a comfortable shower, which is unsurprising. In other words, the agent does not continue to infer the hidden cause of its original suprising observations (i.e., a very hot shower), but rather *intervenes in the world* so as to bring about preferred states that fit his prior expectations about the sorts of sensations he expects to encounter.

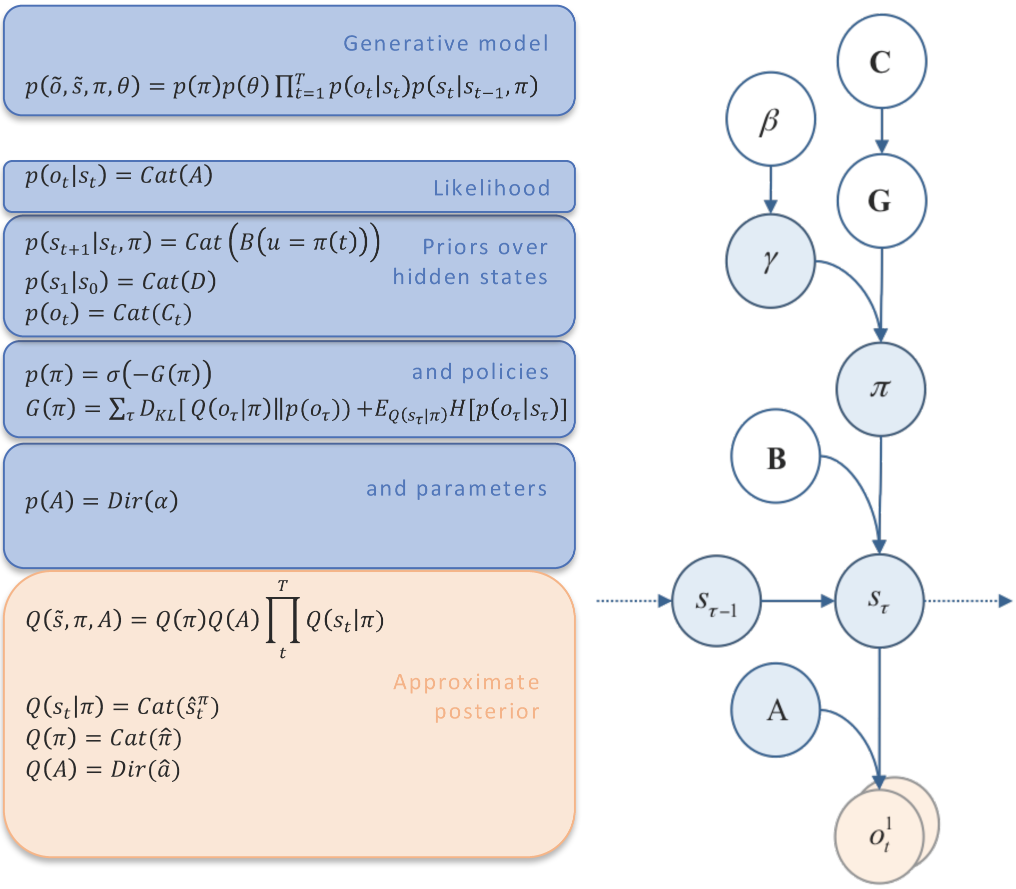
Active inference *with* preferences therefore changes the epistemic pattern the agent engages in. Rather than, analogous to a rigorous scientist, inferring the causal structure of the world by probing it and observing the resulting data, the agent acts like a *crooked* scientist, expecting the world to behave in a particular kind of way and through changing the world, ensures that those expectations come true (Bruineberg et al., 2016 (Chapter 2)).

This changes the interpretation of free-energy minimization: in active inference *without prior* preferences, the minimum of free-energy coincides with an agent that comes to infer the hidden structure of the world, in active inference *with* preferences, the minimum of free-energy is attained when sensations are generated by characteristic or preferred states that are realized through action (Friston, 2011).[4] In this way, crucially, the free-energy principle provides a common currency for both epistemics (finding out about the state of the world) and value (engaging with the world to seek out preferred outcomes). Agents are adaptive if they expect to be in states they characteristically thrive in and, through action, make those expectations come true.

---

[4] If now what the agent prefers is itself a product of its phylogenetic and ontogenetic history, then what results is akin to an enactive theory of cognition (Allen & Friston, 2016; Bruineberg et al., 2016 (Chapter 2)).

What we have shown in this section is that what exactly *is* the minimum of free-energy differs depending on the assumption one makes about the nature of the agent and the task at hand: it coincides with an epistemic fit if one assumes perceptual inference and active inference *without* preferences, and it coincides an epistemically enriched value-based, pragmatic fit in the case of active inference *with* preferences. In the context of certain perceptual decision-making experiments carried out in a lab, such as the widely used random-dot motion task (e.g., Ball & Sekuler, 1982; Newsome & Pare, 1988) it might make sense to treat a rational agent as not having intrinsic preferences for a direction of motion. However, in an ecological setting, what matters is not just what the cause of the current sensory input is, but to be sensitive to the implicit pragmatic and epistemic affordances that enables the selection of actions that lead to preferred, or characteristic, sensory exchanges. Because the prior preferences ensure that creatures act in ways that minimize expected free-energy, if they have the right sort of generative model, agents will, in acting, obtain the sensory evidence they expect. This concludes our formal description of active (embodied) inference and the ensuing sort of self-organisation that emerges from it. We now turn to simulations to illustrate that free-energy minimization cuts both ways in an agent-environment exchange.



Generative model

$$p(\tilde{o},\tilde{s},\pi,\theta) = p(\pi)p(\theta)\prod_{t=1}^{T} p(o_t|s_t)p(s_t|s_{t-1},\pi)$$

Likelihood

$$p(o_t|s_t) = Cat(A)$$

Priors over hidden states

$$p(s_{t+1}|s_t,\pi) = Cat\left(B\left(u = \pi(t)\right)\right)$$
$$p(s_1|s_0) = Cat(D)$$
$$p(o_t) = Cat(C_t)$$

and policies

$$p(\pi) = \sigma\left(-G(\pi)\right)$$
$$G(\pi) = \sum_\tau D_{KL}[\,Q(o_\tau|\pi)\|p(o_\tau))\, + E_{Q(s_\tau|\pi)}H[p(o_\tau|s_\tau)]$$

and parameters

$$p(A) = Dir(\alpha)$$

Approximate posterior

$$Q(\tilde{s},\pi,A) = Q(\pi)Q(A)\prod_{t}^{T} Q(s_t|\pi)$$
$$Q(s_t|\pi) = Cat(\hat{s}_t^\pi)$$
$$Q(\pi) = Cat(\hat{\pi})$$
$$Q(A) = Dir(\hat{a})$$

Figure 5.2: Generative model and (approximate) posterior. *Left panel*: A generative model is the joint probability of outcomes o, hidden states s, policies π and parameters θ: see top equation. The model is expressed in terms of the *likelihood* of an observation o_t given a hidden state s_t, and *priors* over hidden states: see second equation. In Markov decision processes, the likelihood is specified by an array A, parameterized by concentration parameters α (see main text). The empirical priors over hidden states depend on the probability of hidden states at the previous time-step conditioned upon an action u (determined by policies π), these probabilistic transitions are specified by matrix B. The important aspect of this generative model is that the priors over policies $p(\pi)$ are a function of expected free-energy $G(\pi)$. That is to say, *a priori* the agent expects itself to select those policies that minimize expected free-energy $G(\pi)$ (or its path integral $\sum_{\tau} G(\pi, \tau)$). See the main text and 'textbfTable 5.1 for a detailed explanation of the variables. In variational Bayesian inversion, one has to specify the form of an approximate posterior distribution, which is provided in the lower panel. This particular form uses a mean field approximation, in which posterior beliefs are approximated by the product of marginal distributions $Q(s_t \pi)$ over unknown quantities. Here, a mean field approximation is applied to both posterior beliefs at different points in time $Q(s_t \pi)$, policies $Q(\pi)$, parameters $Q(A)$ and precision $Q(\gamma)$. *Right panel:* This Bayesian graph represents the conditional dependencies that constitute the generative model. Blue circles are random variables that need to be inferred, while orange denotes observable outcomes. An arrow between circles denotes a conditional dependency, while the lack of an arrow denotes a conditional independency, which allows the factorization of the generative model, as specified on the left panel.

5.3 Simulation of niche construction

So far, we have addressed the motivation for, and derivation of, the free-energy principle and how actions underwrite the minimization of expected free-energy. We now turn to simulations of niche construction using a free-energy minimizing agent. In order to do this, we need to make specific assumptions about the structure and parameters of the generative model that is constituted by the agent - and the generative process in the econiche. In brief, we will use a very simple model of the world that can be thought of as a maze that can be explored. Crucially, the very act of moving through the maze changes its state; thereby introducing a circular causality between the environment (i.e., maze) and a synthetic creature (i.e., agent), who traverses the environment, in search of some preferred location or goal.

To build this simulation, we will assume some specific conditional independencies that render the generative model a so-called Markov Decision Process (MDP). The main two features of Markov decision processes are i.) that observations at a particular time o_t depend only on the current hidden state s_t, and ii.) the probability of a hidden state s_{t+1} depends only on the previous hidden state s_t and the policy $\pi(t)$ (see **Figure 5.2**, right panel). Each of the probabilistic mappings or transitions is parameterized by a distribution matrix (**Figure**

5.2, left hand side). The outcome or likelihood matrix is given by A, where $A_{ij} = P(o_t = is_t = j)$. The probability transition matrix of hidden states over time is given by B, where $B_{ij}(u) = P(s_{t+1} = i|s_t = j, \pi(t) = u)$. C denotes prior (preferred) beliefs about outcomes $P(o_t)$ and D denotes beliefs about the initial states at t=1. These conditional probabilities can be seen in **Figure 5.2**. As above, we define the variational distribution as:

$$Q(\tilde{s}, \pi, A) = Q(\pi)Q(A)\prod_{t}^{T} Q(s_t|\pi)$$

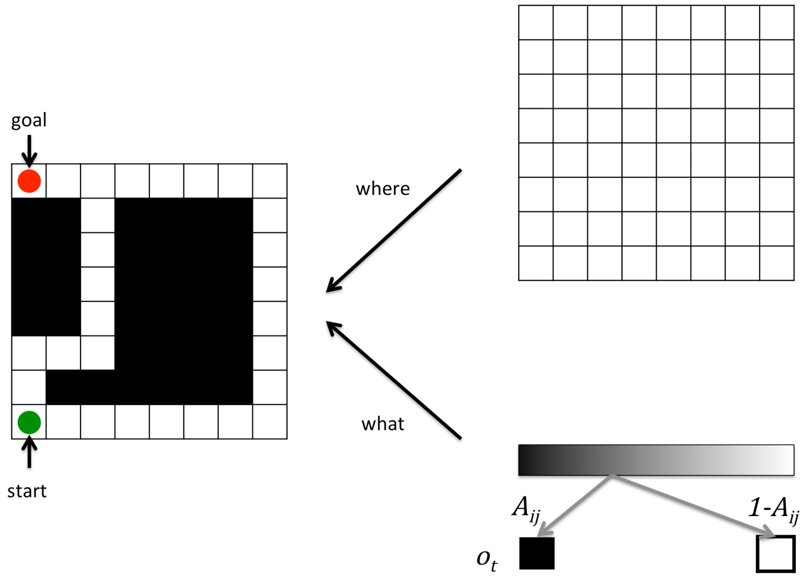


Figure 5.3: The layout of the environment: The agent's environment comprises an 8x8 grid. At each square the agent observes its current location ('where' hidden state) and either an 'open' or 'closed' state ('what' hidden state). The mapping from hidden states to observations in the 'where' modality is direct (i.e., one-to-one). In the 'what' modality, the statistics of the environment are given by the ***A***-matrix. An outcome is generated probabilistically based on the elements of the A-matrix at a particular location. The agent starts at the left bottom corner of the grid (green circle) and needs to go to the left top corner (red circle).

In what follows, we describe the particular form of the generative model - in terms of its parameters, hidden states and policies - that will be used in the remainder of this paper. An agent starts at a specified location (**Figure 5.3** - green circle) on an 8 x 8 grid and is equipped with a prior belief it will reach a goal location (**Figure 5.3** - red circle) within a number of time steps, (preferably) without treading on 'closed' (black) squares. The agent's visual input is limited, in the sense that it can only see whether its current location is open (white) or closed (black). This means that, in the absence of prior knowledge, an agent needs to visit a location in order to gather information about it.

Each trial comprises several epochs. At each epoch, the agent observes its current position, carries out an action: moving up, down, left, right, or stay, and samples its new position. A trial is complete after a pre-specified number of time steps. In addition to visual input, we also equip the agent with positional informational; namely its current location. This means that there are two outcome modalities (o_t): *what* (open/white vs. closed/black) and *where* (one of 64 possible locations) (see **Figure 5.3**). The generative model of these outcomes is simple, the hidden states (s_t): correspond to the 64 positions. The likelihood mapping for the *where*-modality corresponds to an identity matrix, returning the veridical location for each hidden state. For the *what*-modality, the likelihood matrix specifies the probability of observing an open versus a closed state: $A_{ij}^{what} = P(o_t = white|s_t)$, parametrized by concentration parameters (see below). The (empirical) probability transitions are encoded in five matrices (corresponding to the 5 policies of the agent: $B_{ij}^{\pi} = P(s_{t+1} = i|s_t = j, \pi)$. These matrices move the hidden (*where*) states to the appropriate neighboring location given the policy. The D vector designates the true starting location of the agent. Prior beliefs over allowable policies depend on expected free-energy $G(\pi)$, which depends on prior preferences, or costs, over outcomes C (see below). When the parameters are unknown, as is the case for A, the parameters are modeled using Dirichlet distributions over the corresponding model parameters. Based on the particular generative model, one can derive the update equations (**Table 5.2**) that underwrite the minimization of free-energy (see **Appendix B** and Friston, Rigoli, et al., 2015).

Variational updates for the parameters (i.e. expectations) of the approximate posterior distribution

Perception and state-estimation

$$\begin{aligned} s_t^\pi &= \sigma(v_t^\pi) \\ \dot{v}_t^\pi &= \bar{A}\cdot o_t + \overline{B_{t-1}^\pi} s_{t-1}^\pi - \overline{B_t^\pi}\cdot s_{t+1}^\pi - v_t^\pi \\ o_t^\pi &= A\ s_t^\pi \end{aligned}$$

Evaluation and policy selection

$$\begin{aligned} \pi &= \sigma(-F - \gamma\cdot G) \\ F_\pi &= \textstyle\sum_t s_t^\pi\cdot(\ln s_t^\pi - \overline{B_{t-1}^\pi} s_{t-1}^\pi) - \sum_t s_t^\pi\cdot\bar{A}\cdot o_t \\ G_\pi &= \textstyle\sum_t o_t^\pi\cdot(W\cdot s_t^\pi + \ln o_t^\pi + C_t) + H\cdot s_t^\pi \end{aligned}$$

Precision and confidence

$$\begin{aligned} \dot{\hat{\beta}} &= (\pi - \pi_0)\cdot G + \beta - \hat{\beta} \\ \pi_0 &= \sigma(-\gamma\cdot G) \end{aligned}$$

Bayesian model averaging and learning

$$\begin{aligned} E_Q[s_t] &= \textstyle\sum_\pi \pi_\pi \cdot s_t^\pi \\ \ln \widehat{A_t} &= \psi(\alpha) - \psi(\alpha_0) \\ \widehat{a_t} &= a_t + o_t \otimes s_t \end{aligned}$$

Change of the environment

$$\begin{aligned} \ln \widehat{\boldsymbol{A_t}} &= \psi(\boldsymbol{\alpha}) - \psi(\boldsymbol{\alpha_0}) \\ \widehat{\boldsymbol{a_t}} &= \boldsymbol{a_t} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \boldsymbol{s_{t-\mathbb{t}}} \end{aligned}$$

Action selection

$$\boldsymbol{u_{t'}} = \max_u \pi \cdot [\pi(t) = u]$$

Table 5.2: Variational update equations.

5.3.1 Preferred outcomes and prior costs

The problem the agent faces is twofold. First, we want the agent to move from its start location to its target location; however, it can only see its current location and is only able to plan one move ahead. Second, the agent does not like treading on black (closed) squares, but at least initially, does not know which squares are black and which are white. Its job is then to find its way to the target location while avoiding black squares. The A matrix contains the agent's prior beliefs or preferences about outcomes in both modalities - *what* and *where*. At each epoch, the agent updates its prior beliefs based upon what it has come to know about the environment and selects its actions accordingly. In the current simulation, the agent's preferences or prior beliefs are that it will move towards a target location without transgressing into black squares. The subtle issue here is that the agent needs to select a policy that brings it closer to its goal state (taking into account what it knows about the layout of the environment) without performing an exhaustive search or planning into the far ahead future.

Intuitively, the agent's preferences can be understood in the following way: at each epoch, the agent expects to occupy locations that are not black, within the reach of its policies *and* are most easily accessible from the *target* location. Given that the agent's preferences are reconfigured after each epoch, the agent will inevitably end up at its target location. More formally, the expected cost (i.e., negative preference) of a sensory outcome at a future time τ can be described in the following way:

$$\begin{aligned} C_\tau &= -\ln p(o_\tau) \\ &= \ln([\exp(T)s_1 < e^{-3}] + e^{-32}) - \ln \exp(T)s_T \end{aligned}$$

Where:

$$T_{ij} = \begin{cases} -\sum\limits_{i \neq j} T_{ij} & i = j \\ A_i & \exists u : B^u_{ij} > 0 \\ A_i & \text{otherwise} \end{cases}$$

Although the first term might look complicated, it just corresponds to a prior cost (of -32) whenever the condition in square brackets is not met, and zero otherwise. In other words, it assigns a high cost to any location that is occupied with a small probability when starting from the initial location s_1. The second term corresponds to the (negative) log probability a given state is occupied when starting from the target location (s_T), favoring states that are occupied with high probability. Prior beliefs about transitions are encoded in a 'diffusion' matrix $exp(T)$. Formally, T is a graph Laplacian, and the diffusion matrix corresponds to a Green's function. Intuitively this encodes the probability that a state will be occupied following diffusion from any other state during one time step, taking into account the agent's posterior expectations about whether states or open or closed. Specifically, $exp(T)s_1$ and $exp(T)s_T$ designate the probability of diffusion to any location from the starting location and the target location respectively.

The details of this particular prior cost function do not matter too much-they just serve to model preferences that lead to goal-directed behavior under constraints and uncertainty. We have used these priors previously to simulate foraging in mazes (Kaplan and Friston, under review). Here, we use the same setup but generalized to include an effect of navigating through the maze on the maze itself [Matlab code and demo routines detailing this generative model of spatial navigation are available in the **DEM Toolbox** of the **SPM open source software**: http://www.fil.ion.ucl.ac.uk/spm/].

5.3.2 Learning and the likelihood matrix

Although the graph Laplacian provides the agent with prior preferences (i.e., costs C_τ), these are not the only factors underlying policy selection. The expected free-energy also contains an ambiguity term (see above and **Appendix A**) that is minimized when agents minimize the uncertainty of observations afforded by a particular location. This implies that the agent expects to explore its environment, even when this exploration does not bring it closer to its target state. This can be seen in **Figure 5.4**, which shows the results of the simulation of successive trials. In the absence of any accumulated knowledge about the environment, the agent heads straight to its target state and then (rather than stay there) explores the local environment. In the next trial, the agent heads to its target state, while avoiding those locations that it now knows are closed. In the third trial, the agent has found the shortest (open) path to its target state, but still explores its surrounding, whenever in its vicinity ambiguity can be reduced. In trial four, and thereafter, the agent follows its “well trodden” and unambiguous white path.

At the beginning of a series of trials, the agent is initially naive about the structure of the maze. This naivety can be quantified by equipping the agent with priors parameterized by Dirichlet distributions. The underlying concentration parameters of this prior can be thought of as the number of observations (or pseudo-observations) of a particular outcome the agent has already made before the start of a trial. In our case, the agent has separate concentration parameters for each outcome at each location. There are two relevant dimensions for the set of concentration parameters at a particular location: their absolute and their relative size. When the absolute size of the concentration parameters is low, the agent learns the hidden state (open or closed) of a location after one observation. When the concentration parameters - reporting the number of times open or closed outcomes have been experienced - are high, the agent needs many more observations to be convinced a state is open or closed (see **Table 5.3**). In short, the concentration parameters determine both the prior expectations about the world and the confidence placed in those expectations. This confidence or precision determines the impact of further evidence, which decreases with greater confidence.

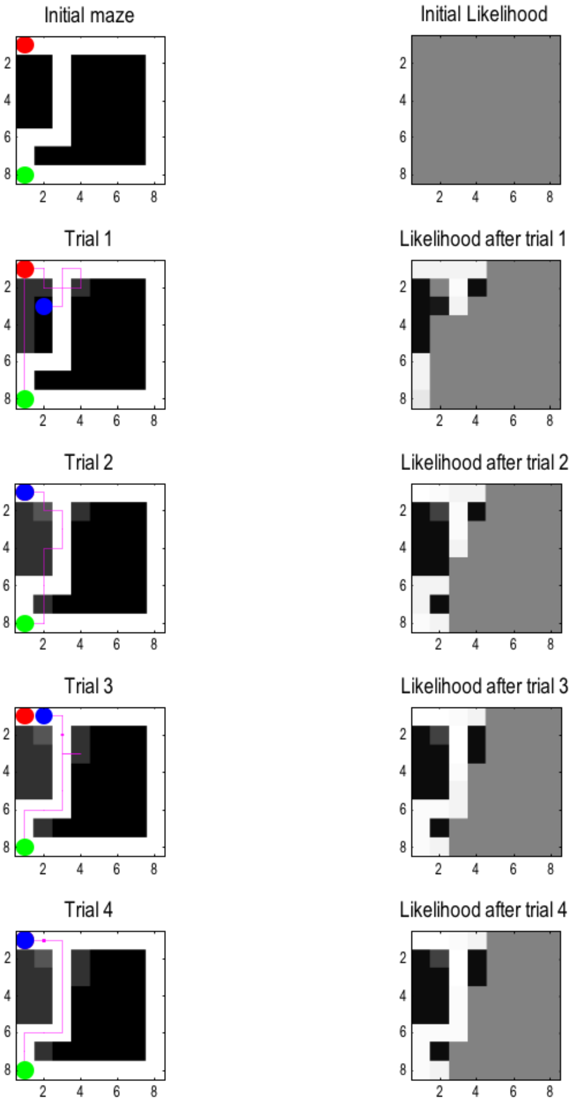


Figure 5.4: Exemplar trials: The left column shows the layout of the environment ($\boldsymbol{A}$-matrix) and the right column shows the agent's expectations about the environment (A-matrix). The rows show the starting condition and the location after each trial. The green, red and blue circles designate the starting, target and final position respectively. The red-dotted line shows the agent's trajectory at other moves within a trial. In this and all subsequent examples, each trial comprised 16 moves. This figure illustrates four consecutive trials and consequent changes in the likelihood matrices that constitute the generative process (i.e., environment) and model (i.e., agent).

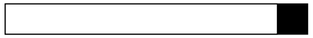
Concentration parameters:	$\alpha_{white} = 0.125$ $\alpha_{black} = 0.125$	$\alpha_{white} = 4.5$ $\alpha_{black} = 4.5$	$\alpha_{white} = 0.5$ $\alpha_{black} = 4.5$
Prior expectation: $E(o_i) = \frac{\alpha_i}{\sum_k \alpha_k}$	$P(o_i) =$ 	$P(o_i) =$ 	$P(o_i) =$ 
Posterior expectation: $E(o_i\|n) = \frac{\alpha_i + n_i}{\sum_k \alpha_k + n_k}$	$P(o_i\|white) =$ 	$P(o_i\|white) =$ 	$P(o_i\|white) =$ 

Table 5.3: Updating of concentration parameters – Prior expectations about the layout of the environment are given by a Dirichlet distribution, which is parameterized by concentration parameters α_{white} and α_{black}. The agent's prior expectation about the state of the environment can be expressed in terms of the (relative value of the) concentration parameters. Concentration parameters are updated in proportion to the number of observations of a particular outcome.

Crucially, different prior settings of the concentration parameters lead to qualitatively different behaviors. In **Figure 5.5** we illustrate the different behaviors the agent exhibits as a function of its initial concentration parameters. This figure shows the trajectories of agents at their fourth trial. The fast-learning, or naive, agent with low concentration parameters (left) finds the route to the target, where its learning history is shown in **Figure 5.4**. An agent with intermediate concentration parameters (middle) needs more observations to learn a particular location is open or closed. Once it is confident enough that the intervening region - between its current location and its target location - is closed, it will stay put in an open location. The slow-learning, or stubborn, agent with high concentration parameters (right) is, after four trials, convinced that the locations it has frequented are closed. In subsequent trials, it will explore a trajectory parallel to its current one, and once it knows these states are also closed, stays put in the same place as the agent with medium concentration parameters. Although all three agents start with the same set of beliefs about the structure of their environment, they each ascribe different levels of confidence to these beliefs. This means that they learn (change these beliefs) at different rates, resulting in qualitatively different behaviors. We will use this simple but fundamental difference among agents or phenotypes to illustrate the remarkable impact these differences in prior beliefs can have on econiche construction in later simulations.

5.3.3 The environment adapting to an agent

So far, we have considered a stationary environment. That is to say, an agent can move around and selectively sample from its environment, but not change it.[5] Things change profoundly when we allow the agent to change the statistical structure of the environment itself. In the following simulations, we parameterized the generative process with a Dirichlet distribution, just as we did for the generative model. In particular, we now have both an observation matrix A, embodying what the agent believes about the mapping between locations s and observations o, and an generative matrix A, denoting the *actual* mapping between locations s and observations o. The update equations for the observation matrix and generative matrix (bold) reflect the implicit symmetry of agent-environment interactions:

---

[5] In fact, strictly speaking, the simulations did allow the environment to change because we used prior concentration parameters of 4 for the environment. One can see this in the upper panels of **Figure 5.5**, which shows the environmental likelihood matrix changes slightly, after four trials or paths.

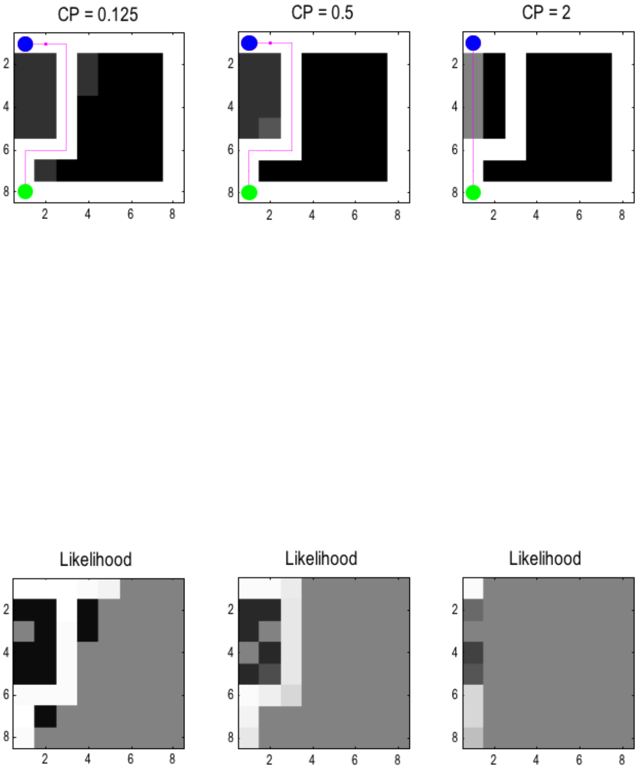


Figure 5.5: Dependency on concentration parameters: The figures show the environment (in terms of the likelihood of outcomes at each location) and trajectories (top) and expectations (bottom) after the 4th trial for agents with prior concentration parameters of 1/8, 1/2, and 2 respectively. The expected likelihood (lower row) reports the agent's expectations about the environment (i.e., the expected probability of an open – white – or closed – black – outcome). We see here that with low priors the agent is more sensitive to the outcomes afforded by interaction with the environment and quickly identifies the shortest path to the target that is allowed by the environment. However, as the agent's prior precision increases, it requires more evidence to update its beliefs; giving the environment a chance to respond to the agent's beliefs and subsequent action. In this case, a 'desire' path (i.e., shortcut) is starting to emerge after just four trials (see upper right panel). We focus on this phenomenon in the next figure.

$$
\begin{aligned}
\widehat{A_t} &= Dir(\widehat{a}_t) \\
\widehat{a}_t &= a_t + o_t \otimes s_t \\
\\
\widehat{\boldsymbol{A_t}} &= Dir(\widehat{\boldsymbol{a_t}}) \\
\widehat{\boldsymbol{a_t}} &= \boldsymbol{a_t} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \boldsymbol{s_t}
\end{aligned}
$$

The concentration parameters a of the observation matrix at time t are updated by adding $+1$ to the concentration parameter of a particular outcome o_t at a particular location s_t. The concentration parameters a of the generative matrix at time t are updated by adding $+1$ to the concentration parameter of the *open* outcome at the location that the agent visited. In other words, the more often an agent visits a particular location, the more likely this location will provide the agent with open observations. The motivation behind these update rules was to show how easily so-called 'desire paths' can emerge: the more a path through long grass is trodden, the more 'walkable' it becomes.

The relative value of the environmental concentration parameters $\boldsymbol{a}$ determines the probability of a particular location providing the agent with an open or closed observation. In all initial situations, we set the concentration parameters to either a low value (1/8) or a high value (1024). The absolute value of the concentration parameters can be interpreted in exactly the same way as in the generative model; namely, the propensity to update in light of new evidence. Here, the evidence is provided by action of the agent on the environment and the propensity for environmental updates corresponds to the *inertia* of the environment, or the ability of the environment to 'remember' the trajectory of the agent. In short, the environment can impress an agent to a greater or lesser extent, depending upon the agent's prior beliefs. In exactly the same way, an environment may be, literally, impressed by an agent - to a greater or lesser extent. The degree of 'impression' in both cases rests upon the prior precisions encoded by (in this example) prior concentration parameters in the generative model (agent) and generative process (environment) respectively.

Figure 5.6 shows the effects of the different prior concentration parameters on the dynamics of both the agent's observation matrix A and the environmental generative matrix $\boldsymbol{A}$. As above, this figure shows the path at the fourth trial, as well as the underlying A and A at the end of the fourth trial. The bottom row is similar to **Figure 5.5**: when the environment has high concentration parameters, the agent takes a very long time to change the statistics of the environment. The upper left panels report the situation where concentration parameters are low for both the agent and the environment. The trajectory of the agent over the four trials is identical to the trajectory of the agent with high environmental concentration parameters (bottom right). Since the agent learns a location is closed at once, it never revisits the location to confirm its beliefs, and will therefore

Low-inertia environment

High-inertia environment

Figure 5.6: Dependency on concentration parameters of the agent and environment: This figure shows the layout of the environment (***A***-matrix) and the agent's expectations about the environment (A-matrix) at the end of the 4th trial, as a function of the prior concentration parameters of both the agent and the environment. The left and right columns show the trajectory for high and low learning rates for the agent (with prior concentration parameters of 1/8 and 2), respectively. The top and bottom row show the trajectory for high and low learning rates of the environment (prior concentration parameters of 1 and 16), respectively. Note the unambiguous emergence of a 'desire' path in all scenarios apart from an environment with high concentration parameters and an agent with low concentration parameters (bottom left); i.e., an agent who is willing to learn but an environment that is not yielding. The most unambiguous desire path is clearly evident when the agent is relatively fastidious (with high prior concentration parameters) and the environment is compliant (with low concentration parameters (upper right)).

not learn about the environmental changes. Although a more efficient path has become available to the agent it is unable to exploit this path because it is places too much confidence in its past experience to explore alternative policies; i.e., its prior beliefs have precluded openness to any epistemic affordance. The upper right panels report the context where concentration parameters are high for the agent, and low for the environment. Like all agents, the agent starts out heading directly for the target state, but in so doing changes the generative matrix $\boldsymbol{A}$ so that it is more likely to provide the agent with open observations. Because the learning rate of the agent is slower than the rate of change of the environment, the agent carves out an open path by moving repeatedly down the same path (without knowing it has done so).

In summary, depending on the prior concentration parameters of both the agent and the environment, the agent either 1.) learns (and consolidates) the initial path through its environment, 2.) learns the initial path through its environment, but, in learning, opens up new paths, 3.) does not learn an open path or 4.) carves out a new path to its target location.

5.3.4 Agent-environment convergence

Over time, the agent learns the structure of its environment while the environment accumulates knowledge about the agent's behavior, which depends - in a circular fashion - on the agents expectations. We can quantify the implicit coupling between the agent and environment by exploiting the symmetry between the generative matrix $\boldsymbol{A}$ and observation matrix A. This symmetry allows us to create a 'phenotypic space' that is shared by the agent and environment; namely, the patterns of concentration parameters (of both the generative and observation matrix) that show the greatest changes over time. These patterns can be obtained efficiently as the principal components or eigenvectors of the covariance matrix between expectations in both matrices over time.

These eigenvectors define a metric space that summarizes expectations about the consequences of being in any particular location. Crucially, this space is shared by the agent and environment, which allows us to plot the evolution of the agent - and the environment - in the same space and ask how they move in relation to one another. Furthermore, we can visualize the influence of the environment on the agent and *vice versa* in a compact form via trajectories in this phenotypic space. We will use the (space spanned by the) first two eigenvectors to depict the coupling between the agent and the environment. For ease of visualization, we used the deviations from the final expectations of the agent and environment for each simulation. This ensures that their respective trajectories converge on the same point in phenotypic space.

Figure 5.7 plots the corresponding trajectories for the agent (black closed circles) and the environment (red open circles) for each of the four conditions (high and low concentration parameters in the agent and the environment respectively).

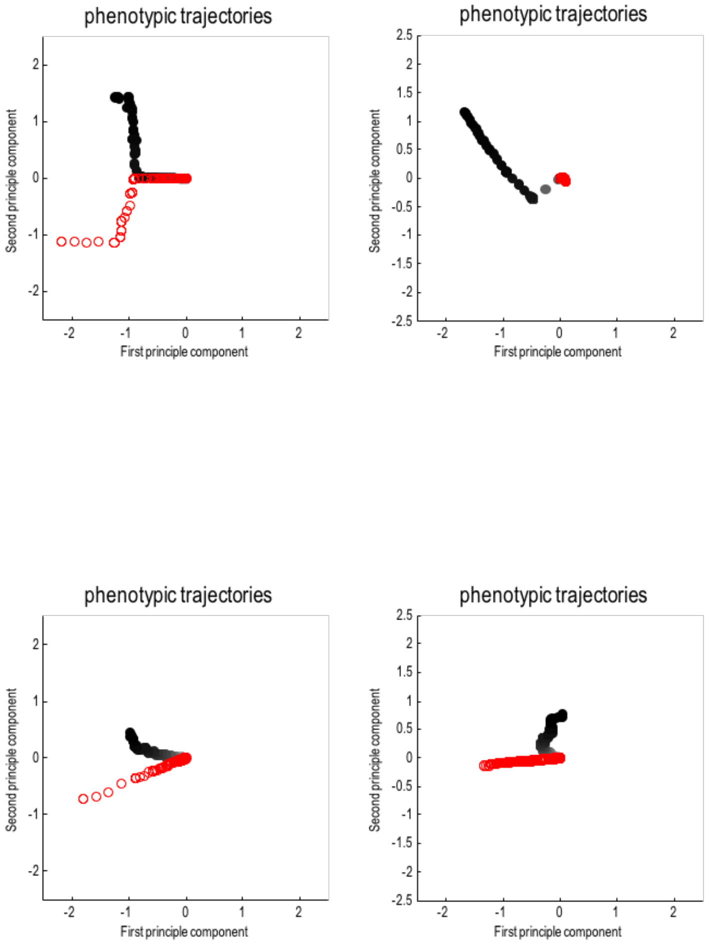


Figure 5.7: Trajectories of agent and environment in phenotypic space: the phenotypic space is defined by the first two eigenvectors of the covariances among the expectations of (both agent and environment) of an open outcome, at each location, over time. The upper and lower panels show the trajectory for low and high prior precision for the agent (with initial concentration parameters of 1/8 and 2), respectively. The left and right panels show the trajectory for low and high prior precision of the environment (with initial concentration parameters of 1 and 16), respectively. Open and closed circles designate the environment and the agent respectively, while the grey scale designates the evolution over time. In this example, the trajectories converge to the same point in phenotypic belief space because the expectations were expressed as deviations from the respective final expectations of the agent and environment.

This licenses a metric interpretation of how the agent's expectations evolve over time (the learning rate), the changes in environmental expectations (the inertia) and the movement of both the agent's expectations and the environment, with respect to each other. The upper and lower panels show the trajectory for low and high prior precision for the agent (with initial concentration parameters of 1/8 and 4), respectively. The left and right panels show the trajectory for low and high prior precision of the environment (with initial concentration parameters of 1 and 16), respectively. Open and closed circles designate the environment and the agent respectively, while the grey scale designates the evolution over time.

The key thing to take from these results is the relative excursion of the environment and agent in their shared phenotypic spaces. It is immediately apparent that the relative prior precision of (implicit) beliefs held by the agentm and environment determine how much they move in this space. For example, when both have a low prior precision (in terms of concentration parameters) both move substantially through phenotypic space and crucially, converge on the same direction after a sufficient period of time (see upper left panel). What is remarkable here is that the direction through phenotypic space coincides when the environment and agent are sensitive to each other. Conversely, when the environment is less responsive (i.e., has higher concentration parameters) it moves relatively little in the phenotypic space - while the agent does all the heavy lifting in terms of adapting to the environment. The lower panels show the equivalent excursions when the agent has greater convictions in its prior beliefs (with high prior concentration parameters). The key thing to observe here is that large distances are traversed in phenotypic space and there is a failure to find a common direction.

This example illustrates an interesting and possibly counterintuitive phenomenon; namely, that the learning 'about each other' depends in a sensitive way on the relative confidence placed in prior beliefs (i.e., the Dirichlet parameters in this example). This confidence has a profound effect on the rate at which the agent learns about the environment and *vice versa* - and the degree to which their respective expectations converge.

5.3.5 Fitness and performance

Using the principal components (i.e., eigenvectors) to define a joint phenotypic space allows one to visualize the development or learning trajectories of the agent and environment. However, this does not mean that the metric distance in phenotypic space reflects the 'fitness' of the agent-environment system. In terms of fitness, what matters is the time integral of variational free-energy (free action), given the locations that are actually visited (and outcomes experienced). More formally, from the perspective of the free-energy principle fitness corresponds to model evidence (Campbell, 2016; Frank, 2012):

$$\ln P(\tilde{o}) \approx -F(\tilde{s}, \pi, \theta_i)$$

In other words, the model evidence is scored by the minimum of free-energy, given a set of observations $\tilde{o}$ and a set of parameters θ_i (most notably the A-matrix learned after a series of trials). On this interpretation of free-energy, one could assess the 'fitness' of a range of agents (parameterized by θ_i) for a given set of environmental data $\tilde{o}$. However, the point of this paper is not to interpret evolution and learning as the optimization of parameters given a fixed environment, but rather *as the convergence of both agent and environment as a function of their reciprocal interaction.* The environmental data $\tilde{o}$ is itself dependent on the statistics of the environment $\boldsymbol{\theta_j}$ (specific to a context j) and the policies the agent pursues:

$$F_{i,j} = \min_{\tilde{s},\pi} F(\tilde{s}, \pi | \tilde{o}_{i,j}, \theta_i)$$

Here, $\tilde{o}_{ij}$ is the sensory data received by agent i in environment j.

This allows us to evaluate and compare the fitness of each of the four agents in each of the four environments. For simplicity, we have disabled learning of the agent as well as adaptation of the environment. The result is a 4x4 matrix that rates the accumulated free-energy for a trial for an agent i in environment j.

Figure 5.8 shows the changes in variational free-energy, expected free-energy and reaction times (i.e., computational time taken to execute a path) during 32 successive exposures to the environment, where each exposure or path comprised 16 moves. These are the same simulations reported in **Figure 5.7**. The solid lines report the changes in free energies and reaction times for agents with low prior concentration parameters or confidence in their beliefs about (location-dependent) outcomes. These can be thought of as relatively naive agents who have little experience of any world. Conversely, the dotted lines refer to agents who are a more experienced (with prior concentration parameters of 4). As above, these two sorts of agents were exposed to environments that were malleable (black lines) and less sensitive to the agent's behavior (red lines), in virtue of being equipped with prior concentration parameters of 1 and 16 respectively. There are a number of interesting behaviors that these results feature.

First, notice that the free energies fluctuate from exposure to exposure. This reflects the fact that the free-energy is a function of sensory encounters that change from moment to moment. The second, perhaps counterintuitive, observation is that the (negative) variational free-energy *decreases* on the first few trials. Note that we have plotted negative free-energy in the graphs, which can be interpreted as the quality or 'fitness' of exchange with the environment. This initial decrease in free-energy reflects the fact that the environment is changing and the agent's model is playing "catch up". The key thing here is that the free-energy becomes relatively stationary as the agent and environment "get to know each other". This captures the essence of the variational principle of least of stationary action that underwrites the (nonequilibrium) steady-state that active inference aspires to.

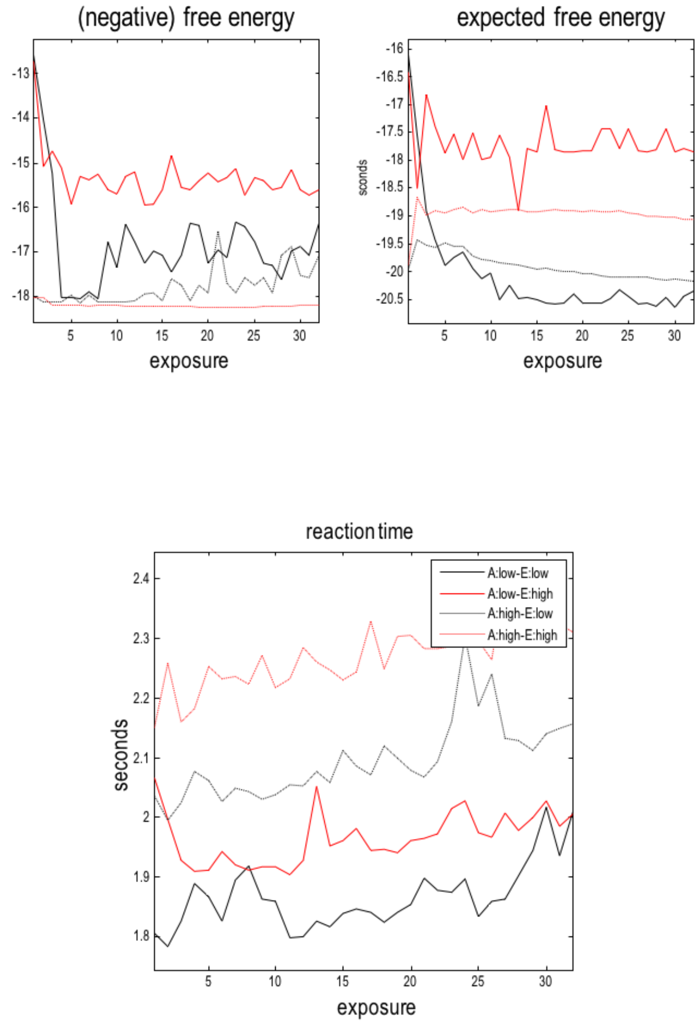


Figure 5.8: Evolution of free-energy: these graphs report the progressive changes in (negative) variational and expected free-energy (upper panels) and simulated reaction times (lower panel) averaged over 16 moves of 32 successive exposures to the environment. The results are shown for an agent with low (solid lines) and high (dotted lines) prior concentration parameters or confidence in its beliefs – in environments with low (black lines) and high (red lines) prior concentration parameters (red lines).

Third, there are some interesting differences between the four simulations. The naive or inexperienced agent in a rigid (high prior concentration parameter) environment appears to fare best - in terms of having the lowest free-energy. In other words, the naive agent learns quickly about its unambiguous world and diligently follows the path specified by environmental cues. It therefore avoids all the uncertainty and ambiguity about having to choose between potential shortcuts and the path evidenced by the environment. However, this is not the case for the naive agent in a malleable environment. Here, the environment itself changes as a result of being explored, which means that the agent's generative model is never quite fit for purpose. Although this agent quickly carves out a shortcut, there is a price to be paid in terms of the uncertainty about what will be observed (and what the best course of action is). Note how the black line dips sharply (in the upper left panel) before recovering to steady-state free-energy levels.

The more cautious agents (dotted lines) show a different sort of dissociation in terms of free-energy. The cautious agent - in a malleable environment - takes a little longer to carve out its shortcut and subsequently learn the consequences of the impressions it leaves on the environment. This results in a slow but progressive decrease in free-energy, in contrast to the same sort of agent in a rigid environment - that never quite offers an unambiguous shortcut. As a consequence, the agent is persistently and mildly surprised by the outcomes it encounters. The evolution of expected free-energy (shown as negative expected free-energy in the figure) follows the same sort of trend. Again, perhaps counterintuitively, the naive agent in a rigid environment appears to be the 'happiest' - in the sense of expecting the lowest free-energy, while the naive agent in a compliant environment always expects to be mildly surprised, in virtue of the fact that it keeps changing the environment it is trying to predict.

Finally, the reaction times (i.e., the computational times averaged over all moves that constitute a path) show two interesting features. First, there is a generic increase in computation time with experience. This reflects the fact that the agent's generative model is becoming more precise as it requires experience. The resulting increase in prior precision translates into an increase in complexity and computational cost. This relationship between precision and computational complexity (i.e., reaction time) is mirrored in terms of the differences among the different simulations, with experienced agents expressing the longest reaction times - and environments with greater prior precision appearing to supplement this computational cost. Clearly, these are anecdotal observations; however, they speak to the interesting relationship between the dynamics of perception and the probabilistic fundaments of active inference.

5.4 Conclusion

To summarize, we have presented an active inference scheme that exhibits epistemic foraging, goal-directed behavior and (unintentional) niche construction using a minimal setup. The key contribution of this paper is to show that *free-energy minimization is a process of the mutual adaptation of agent and environment*: the agent learns from the environment by exploration and the agent's exploration changes the environment until attracting set of states in the agent-environment system is attained. One should note the formal similarity between the update equations for the environment ($\boldsymbol{A}$-matrix) and for the agent (A-matrix) used in this paper. Each is parameterized in terms of the underlying concentration parameters of a Dirichlet distribution, and both the agent and the environment 'accumulate concentration parameters' at places the agent frequents. Formally speaking, this means that the environment 'infers' or 'remembers' the expectations of the agent in the same way as the agent 'infers' or 'remembers' the layout of the environment. What matters from the perspective of the free-energy principle is the convergence of the agent and environment to a free-energy minimum - that is only defined for a particular agent in a particular environment.

Of course, the agent and environment are not completely symmetric: in the current simulations, the environment is fairly simple and is merely reactive; it does not form expectations about the behavior of the agent and does not tend to optimize itself by luring the agent into particular behaviors. However, it is not hard to imagine more active niches, for example environments populated with other agents. One can think of an environment consisting of multiple agents, where the sensory states of one agent are generated by the action of the other agents. Over time, the agents mutually constrain each other until an attracting (synchronization) manifold is reached (Friston & Frith, 2015). In such a case, a stubborn agent (one with high concentration parameters) might persist in its behavior despite evidence to the contrary. In so doing, it forces more flexible agents (with lower concentration parameters - or less confidence in their prior beliefs) to adapt to the behavior of the confident agent. This makes the behavior of the confident agent the predominant determinant or 'driver' of joint dynamics. This circular causality between an agent and its environment will be an important avenue for future research.

Note, finally, that we *could* have equipped the agent with knowledge about how its own actions change the statistics of the environment. This could be done by equipping the agent with beliefs that a change in the A-matrix depends on its action. This would lead to intentional niche construction; a behavior in which agents plan the best route through the environment and then carve out that route. In the present context, this would be less interesting, because everything we want to show (the emergence of adaptive shortcuts or desire paths in the environment), would already be provided to the agent. By not equipping the agent with this knowledge, we can investigate niche construction that emerges from the agent's epistemic foraging and goal-directed behavior, rather than as the result of internal planning.

In conclusion, this paper offers a proof-of-principle simulation of niche construction under the free-energy principle. The key point of this paper is that the minimum of free-energy is not at a point in which the agent is maximally adapted to the statistics of a static environment, but can better be conceptualized as an attracting manifold within the joint agent-environment state-space as a whole, which the system tends toward through mutual interaction.

Appendix A: Free-energy

We have defined free-energy in terms of a generative model $P(\tilde{o}, \tilde{s}, \pi, \theta)$ and an arbitrary variational distribution $Q(\tilde{s}, \pi, \theta)$. The free-energy can be written in several forms to show what its minimization entails:

$$F(\tilde{s}, \pi, \vartheta) = \underbrace{D_{KL}[Q(\tilde{s}, \pi, \theta) \| P(\tilde{s}, \pi, \theta | \tilde{o})]}_{divergence} - \underbrace{\ln P(\tilde{o})}_{\log evidence}$$

Optimizing the variational distribution $Q(\tilde{s}, \pi, \theta)$, to minimize free-energy implies that the divergence between the variational distribution $Q(\tilde{s}, \pi, \theta)$ and the posterior $P(\tilde{s}, \pi, \theta | \tilde{o})$ is minimized, rendering $Q(\tilde{s}, \pi, \theta)$ an approximate posterior. Furthermore, because the KL-divergence is always greater than zero, minimizing free-energy provides an upper bound on the negative log evidence upper bound on the negative log evidence.

$$F(\tilde{s}, \pi, \theta) = \underbrace{D_{KL}[Q(\tilde{s}, \pi, \theta) \| P(\tilde{s}, \pi, \theta)]}_{complexity} - \underbrace{E_Q[\ln P(\tilde{o} | \tilde{s}, \pi, \theta)]}_{accuracy}$$

This formulation shows that free-energy is a trade-off between complexity (defined as the divergence between the variational distribution $Q(\tilde{s}, \pi, \theta)$ and the prior $P(\tilde{s}, \pi, \theta)$) and accuracy (defined as surprise of observations under the variational distribution).

$$F(\tilde{s}, \pi, \theta) = \underbrace{-E_{Q(\tilde{s}, \pi, \theta)} \ln P(\tilde{o}, \tilde{s}, \pi, \theta)}_{energy} - \underbrace{H[Q(\tilde{s}, \pi, \theta)]}_{entropy}$$

This formulation shows the analogy between variational free-energy and Helmholtz free-energy in thermodynamics. It also shows that the free-energy can be expressed in terms of two quantities that the agent has access to: namely, the (sufficient statistics) of the the variational distribution and a generative model.

The generative model is defined as:

$$P(\tilde{o}, \tilde{s}, \pi, \theta) = P(\pi | \theta) P(o_1 | s_1, \theta) P(s_1 | \theta) P(\theta) \prod_{t=2}^{T} P(o_t | s_t, \theta) P(s_t | s_{t-1}, \pi, \theta)$$

Where $P(\theta)$ denotes prior probabilities over model parameters, and $P(o_t | s_t, \theta)$ and $P(s_t | s_{t-1}, \pi, \theta)$ denote a likelihood matrix and probability transition matrix respectively. The outstanding question is how the prior over policies $P(\pi | \theta)$ is to be defined.

The logic here is that if an agent expects itself to follow policies that lead to adverse outcomes, the agent would quickly cease to exist. Any agent that does not cease to exist would therefore expect itself to follow policies that it expects to minimize free-energy. This can be expressed by making the prior over policies a Gibbs or softmax function of (negative) expected free-energy G:

$$p(\pi|\theta) = \sigma(-G(\pi))$$

Expected free-energy

The expected free-energy for a particular policy is the energy of counterfactual observations and hidden states expected under their posterior predictive distribution $Q(o_\tau, s_\tau|\pi)$ minus the entropy of the posterior predictive distribution of the hidden states:

$$G(\pi) = \sum_{\tau} G(\pi, \tau)$$

$$G(\pi, \tau) = \underbrace{-E_{Q(o_\tau, s_\tau|\pi)}[\ln P(o_\tau, s_\tau|\pi, \theta)]}_{energy} - \underbrace{H[Q(s_\tau|\pi)]}_{entropy}$$

where $\tilde{Q} = Q(o_\tau, s_\tau|\pi) = P(o_\tau|s_\tau)Q(s_\tau|\pi)$. In other words, the expectation Q is over hidden states and outcomes that will be observed in the future (and not over hidden states and policies, as was the case for the variational free-energy). Intuitively, this can be thought of as the free-energy one expects in the future, if one were to pursue a particular policy.

One can express expected free-energy in a number of ways: assuming $P(o_\tau, s_\tau|\pi, \theta) \approx Q(s_\tau|o_\tau, \pi, \theta)P(o_\tau)$, we have (see **Appendix A** of Friston, Rigoli, et al. (2015) for a derivation):

$$G(\pi, \tau) = \underbrace{D_{KL}[Q(o_\tau|\pi)\|P(o_\tau)]}_{expected\ cost} + \underbrace{E_Q[H[P(o_\tau s_\tau)]]}_{expected\ ambiguity}$$

This expression mean that minimization of $G(\pi, \tau)$ entails minimizing the KL-divergence between (prior) preferred observations and the expected observations under a particular policy (i.e., expected cost) - and minimizing the expected entropy of an outcome under a particular policy (i.e., expected ambiguity). Hence, policies are considered more likely if they realize prior preferences while, at the same time, avoiding ambiguous outcomes that can resolve uncertainty about the hidden or latent states of the world.

Appendix B: Update equations

We have parameterized the generative model as follows:

$$P(\tilde{o}, \tilde{s}, \pi, A) = P(\pi)P(A)P(s_1)P(o_1|s_1, A)\prod_{t=2}^{T} P(o_t s_t, A)P(s_t|s_{t-1}, \pi)$$

and, using the mean field approximation, we have defined our variational distribution as:

$$Q(\tilde{s}, \pi, A) = Q(\pi)Q(A)\prod_{t}^{T} Q(s_t|\pi)$$

If we take the following definition of free-energy:

$$F(\tilde{s}, \pi, A) = \underbrace{D_{KL}[Q(\tilde{s}, \pi, A)\|P(\tilde{s}, \pi, A)]}_{complexity} - \underbrace{E_Q[\ln P(\tilde{o}|\tilde{s}, \pi, A)]}_{accuracy}$$

we can now decompose the free-energy using the conditional independencies in the variational distribution and the generative model:

$$F(\tilde{s}, \pi, A) = \sum_t E_Q[F(\pi, t)] + D_{KL}[Q(\pi)||P(\pi)] + D_{KL}[Q(A)||P(A)]$$

Where:

$$F(\pi, t) = \underbrace{D_{KL}[Q(s_t|\pi)||P(s_t s_{t-1}, \pi)]}_{complexity} - \underbrace{E_Q[\ln P(o_t|s_t)]}_{accuracy}$$

Using the facts that $p(\pi) = \sigma(-G(\pi))$, and $D_{KL}[Q(x)||P(x)] = E_Q[\ln Q(x) - \ln P(x)]$, we can write this as:

$$F(s, \pi, A, \gamma) = E_{Q(\pi)}[F(\pi, t) + \ln \pi - \gamma \cdot G] + E_{Q(A)}[\ln Q(A) - \ln P(A)] + \cdots$$

We can now minimize the free-energy F by finding the functional derivates of F with respect to all of the elements of the variational distribution $Q(s, \pi, \theta)$ and equate them to 0, under the constraint that each of the elements of Q expresses a probability distribution (i.e., sums up to one). This is naturally done using Lagrange multipliers (Beal, 2003; Friston, Trujillo-Barreto, & Daunizeau, 2008). When for example calculating the derivative of F with respect to $Q(s_{t'}|\pi)$, we find:

$$\tilde{F} = F - \lambda\left(\sum_s Q(s_{t'}|\pi) - 1\right)$$

$$\frac{\partial \tilde{F}}{\partial Q(s_{t'}|\pi)} = 0$$

Plugging in the expression for F, we find:

$$-E_{/Q(s'_t\pi)} \ln P(o, s, \pi, \theta) + \ln Q(s_{t'}|\pi) - 1 - \lambda = 0$$

where $E_{/Q(s'_t\pi)}$ designates the expectation with respect to all factors of Q except $Q(s_{t'}|\pi)$. Rearranging and combining all terms not dependent on $Q(s_{t'}|\pi)$ in a constant $\ln Z$, we find:

$$\begin{aligned} \ln Q(s_{t'}|\pi) = \quad & E_{/Q(s'_t\pi)}[\ln P(o_{t'}|s_{t'}) + \ln P(s_{t'}|s_{t'-1},\pi) + \ln P(s_{t'+1}|s_{t'},\pi)] \\ & - \ln Z \end{aligned}$$

Since the only terms that depend on $Q(s_{t'}|\pi)$ come from its Markov blanket, we can write this as:

$$\begin{aligned} \ln Q(s_{t'}|\pi) = \quad & \ln P(o_{t'}s_{t'}) + E_{Q(s_{t'}-1)} \ln P(s_{t'}|s_{t'-1},\pi) \\ & + E_{Q(s'_t+1)} \ln P(s_{t'+1}|s_{t'},\pi) - \ln Z \end{aligned}$$

The transition probabilities $P(o_t|s_t)$ and $P(s_t|s_{t-1},\pi)$ can be expressed using the A and B matrices of the generative model. In order to ensure that free-energy is minimized and inference settles on the belief specified by the equation above, we can simply define the change of current belief $\dot{s}^{\pi}_{t'}$ as proportional to difference between our current belief $s^{\pi}_{t'}$ and the free-energy minimizing belief specified above. The resulting dynamics then perform a gradient descent on free-energy - to settle on the beliefs that minimize free-energy.

$$\begin{aligned} s^{\pi}_{t'} \quad & = \sigma(v^{\pi}_{t'}) \\ \dot{v}^{\pi}_{t'} \quad & = o_{t'} \cdot \overline{A} + \overline{B^{\pi}_{t'-1}} s^{\pi}_{t'-1} - \overline{B^{\pi}_{t'}} \cdot s^{\pi}_{t'+1} - v^{\pi}_{t'} \end{aligned}$$

These are the variational update equations as denoted in **Table 5.2** and have a degree of biological plausibility in the sense that they are ordinary differential equations. The remaining belief updating the other equations can be derived in an analogous manner. Note that while the free-energy is minimized separately for each factor of $Q(s_{t'}|\pi)$, the free-energy depends on its Markov blanket $Q(s_{t'-1}|\pi)$, $Q(s_{t'+1}|\pi)$, and observations $o_{t'}$, which are themselves minimized. The resulting message passing scheme comprises a series of coupled differential equations that, at each time step $t \rightarrow t+1$, is perturbed by an observation o_{t+1}. Within that time step, the system relaxes to a new fixed point. By construction, the specifics of the differential equations $\dot{v}^{\pi}_{t}$ ensures that the fixed point coincides with the minimum of free-energy. Although it is not the main focus of the current paper, such update equations can be linked to hierarchical message passing in the brain (Friston et al., 2017).

Conclusion

> One of the comparatively neglected tasks of a molar environmental psychology is to find out the extent to which environmental hierarchies of probabilities of object-cue as well as of means-end relationships do find a counterpart in similar hierarchies of evaluation by the organism. This would mean that the environmental probabilities be first ascertained for all of the cues or means involved, with say, the 'normal' life conditions taken as the defining reference class (Brunswik, 1943, p. 59).

In this dissertation, I have proposed a conceptual framework for skilled intentionality which drew on three contemporary paradigms: radical embodied cognitive science, ecological psychology and the free-energy principle. This provides a unified account of the different aspects of skilled intentionality (the socio-material, the normative, the dynamical and the phenomenological) and how they relate to one another.

What is central to the account on offer, is an improved understanding of the mutuality between the structure of the agent and the structure of the environment in skilled action. The mutuality between an agent and its environment is characterized in **Chapter 1** of this dissertation as a "landscape of affordances" on the side of the ecological niche, and self-organizing states of action-readiness on the side of the agent. Together, they constitute a selective openness to affordances, by which the agent tends towards grip on a field of affordances (see **Figure 1.1**).

One of the most promising aspects of the free-energy principle, is the integration of value and epistemics under one unifying principle. While absent in rationalist philosophy of mind, this interplay between value and epistemics is at the heart of Merleau-Ponty's notion of the tendency towards an optimal grip (see quote on p. 43). Even when sticking to the rationalist vocabulary, the apt metaphor here is that of a *crooked* scientist: one that takes its own conditions of flourishing as a standard for the way the world *ought* to be. This metaphor and the motivation for it is presented in **Chapter 2**. A comparison between the ecological-enactive approach and a cybernetic and Helmholtzian approach is provided in **Chapter 3**.

What is left implicit in **Chapter 1** is that openness to affordances is made possible by sensitivity to the regularities in the environment. The insistence, by early proponents of ecological psychology, on lawful informational specification of affordances has had adverse consequences on the development of the field: it limits the scope of affordances considerably and fails to notice the contribution of the agent to the perception of affordances. In **Chapter 4** I gave a characterization of environmental regularities in terms of constraints, information and the form of life or ecological niche. The novel account of *general ecological information* allows the affordance framework to be applicable to what is traditionally called 'higher' cognition.

The resulting view is that direct perception and, more generally, responsiveness to affordances requires a symmetry between the structure agent and the structure of the environment. **Chapter 5** shows how this symmetry follows directly from the free-energy principle. It also offers a proof-of-principle simulation of the mutual adaptation of agent (a generative model) and environment (a generative process) under the free-energy principle.

The epigraph of this conclusion by *ecological psychologist* Egon Brunswik (1943), a much-neglected predecessor of Gibson, exemplifies in a beautiful way some of the main findings of this dissertation. The regularities in the environment and the way they point to affordances and place-affordances (general ecological information) needs to be matched on the side of the agent with a similar dynamical and hierarchical structure. Only then can direct perception take place. But in order for responsiveness to affordances to be adaptive, the agent needs to take its own conditions of flourishing (what Brunswik calls the 'normal' life conditions) as its standard. This is the essence of the 'crooked scientist' argument as developed in **Chapter 2**.

I hope to have made clear that what is on offer is both a scientific research program into the structures that enable our selective sensitivity to the affordances offered by the environment, as well as a comprehensive philosophical world-view in which we can be in direct contact with the aspects of the environment that matter to us. Although this dissertation focusses on skilled intentionality as a means by which an agent can flourish, skilled intentionality can have its darker aspects as well. Because intentionality is not something internal to the head (and insulated from the world), but rather an intrinsic aspect of brain-body-environment systems, it is fragile as well. There is a thin line between 'education of attention' and 'manipulation of attention', where the latter does not benefit our own flourishing but rather exploits our attention for other gains (Wu, 2017; Eyal, 2014). The effects of our rapidly changing econiche on our agency and intentionality are only starting to be understood, but I believe the conceptual framework on offer here is both able to capture these ‘vulnerabilities of intentionality’ and show why it matters.

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: MIT Press.

Anderson, M. L. (2017). Of Bayes and bullets. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing* (chap. 4). Frankfurt am Main: MIND Group. doi: 10.15502/9783958573055

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*(03), 204–205.

Apps, M. A., & Tsakiris, M. (2014). The free-energy self: a predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, *41*, 85–97.

Ashby, W. (1952). *Design for a brain.* London, UK: Chapman & Hall.

Ashby, W. (1956). *An introduction to cybernetics.* London, UK: Chapman & Hall.

Auletta, G. (2013). Information and metabolism in bacterial chemotaxis. *Entropy*, *15*(1), 311–326.

Baker, M. D., & Stock, J. B. (2007). Signal transduction: networks and integrated circuits in bacterial cognition. *Current Biology*, *17*(23), R1021–R1024.

Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science*, *218*(4573), 697–698.

Barandiaran, X., & Moreno, A. (2006). On what makes certain dynamical systems cognitive: A minimally cognitive organization program. *Adaptive Behavior*, *14*(2), 171–185.

Barker, R. G. (1968). *Ecological psychology: Concepts and methods for studying the environment of human behavior.* Stanford, CA: Stanford University Press.

Barwise, J., & Perry, J. (1981). Situations and attitudes. *The Journal of Philosophy*, *78*(11), 668–691.

Barwise, J., & Perry, J. (1983). *Situations and attitudes.* Cambridge, MA: MIT Press.

Beal, M. J. (2003). *Variational algorithms for approximate Bayesian inference.* London: University of London.

Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science. *Cognitive Science*, *22*(3), 295–318.

Beer, R. D., & Chiel, H. J. (1993). Simulations of cockroach locomotion and escape. In e. a. Beer R. (Ed.), *Proceedings of the workshop on locomotion control in legged invertebrates on biological neural networks in invertebrate neuroethology and robotics* (pp. 267–285).

Bénard, H. (1900). Les tourbillons cellulaires dans une nappe liquide [the cellular vortices in a liquid layer]. *Rev Gén Sci Pure Appl*, *11*, 1261–1271.

Bernard, C. (1865/1927). *An introduction to the study of experimental medicine.* Courier Corporation.

Bishop, R. C. (2008). Downward causation in fluid convection. *Synthese*, *160*(2), 229–248.

Bitzer, S., Bruineberg, J., & Kiebel, S. J. (2015). A Bayesian attractor model for perceptual decision making. *PLoS Comput Biol*, *11*(8), e1004442.

Briggman, K. L., & Kristan Jr, W. (2008). Multifunctional pattern-generating circuits. *Annu. Rev. Neurosci.*, *31*, 271–294.

Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence*, *47*(1-3), 139–159.

Bruineberg, J. (2017 (Chapter 3)). Active inference and the primacy of the 'I can'. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing* (chap. 5). Frankfurt am Main: MIND Group. doi: 10.15502/9783958573062

Bruineberg, J., Kiverstein, J., & Rietveld, E. (2016 (Chapter 2)). The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. *Synthese*, 1–28.

Bruineberg, J., & Rietveld, E. (2014 (Chapter 1)). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in human neuroscience*, *8*, 599.

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological review*, *50*(3), 255.

Byrge, L., Sporns, O., & Smith, L. B. (2014). Developmental process emerges from extended brain–body–behavior networks. *Trends in cognitive sciences*, *18*(8), 395–403.

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

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

Chemero, A. (2009). *Radical embodied cognitive science.* 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.

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 of London 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.

Clark, A. (1997). *Being there: Putting brain, body, and world together again.* Cambridge, MA: MIT 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. In T. K. Metzinger & J. M. Windt (Eds.), *Open mind* (chap. 7(R)). Frankfurt am Main: MIND Group. doi: 10.15502/9783958570979

Clark, A. (2015b). Radical predictive processing. *The Southern Journal of Philosophy*, *53*(S1), 3–27.

Clark, A. (2016a). Busting out: Predictive brains, embodied minds, and the puzzle of the evidentiary veil. *Nous*, *51*(4), 727-753.

Clark, A. (2016b). *Surfing uncertainty: Prediction, action, and the embodied mind.* Oxford: Oxford University Press.

Clark, A., & Toribio, J. (1994). Doing without representing? *Synthese*, *101*(3), 401–431.

Colombetti, G. (2014). *The feeling body: Affective science meets the enactive mind.* Cambridge, MA: MIT Press.

Costall, A. (1999). An iconoclast's triptych: Edward Reed's ecological philosophy. *Theory & Psychology*, *9*(3), 411–416.

Costall, A. (2004). From Darwin to Watson (and cognitivism) and back again: The principle of animal-environment mutuality. *Behavior and Philosophy*, 179–195.

Dale, R., Dietrich, E., & Chemero, A. (2009). Explanatory pluralism in cognitive science. *Cognitive science*, *33*(5), 739–742.

Darwin, C. (1881). *The formation of vegetable mould, through the action of worms, with observations on their habits.* Mclean: IndyPublish.

Davids, K., Araújo, D., Hristovski, R., Passos, P., & Chow, J. Y. (2012). Ecological dynamics and motor learning design in sport. In N. H. M. Williams (Ed.), *Skill acquisition in sport: Research, theory and practice, 2nd edn.* (pp. 112–130). London: Routledge.

Davidson, D. (1963). Actions, reasons, and causes. *The journal of philosophy*, *60*(23), 685–700.

Davidson, D. (1970/1980). Mental events. In *Essays on actopms and events.* Oxford: Clarendon Press.

Davidson, D. (1999). The emergence of thought. *Erkenntnis*, *51*(1), 511–521.

Dayan, P., & Hinton, G. E. (1996). Varieties of Helmholtz machine. *Neural Networks*, *9*(8), 1385–1403.

De Haan, S. (n.d.). *Enactive psychiatry.* (Forthcoming at Cambridge 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.

De Jaegher, H., & Di Paolo, E. A. (2007). Participatory sense-making. *Phenomenology and the cognitive sciences*, *6*(4), 485–507.

Denys, D., Mantione, M., Figee, M., van den Munckhof, P., Koerselman, F., Westenberg, H., … Schuurman, R. (2010). Deep brain stimulation of the nucleus accumbens for treatment-refractory obsessive-compulsive disorder. *Archives of general psychiatry*, *67*(10), 1061–1068.

Descartes, R. (1633/2011). *De mens.* Amsterdam: Boom.

Descartes, R. (1637/1979). *Over de methode.* Amsterdam: Boom.

Descartes, R. (1641/1988). *Meditations on first philosophy.* Cambridge: Cambridge University Press. (translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch)

Dewey, J. (1976). *Lectures on psychological and political ethics: 1898* (D. Koch, Ed.).

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the cognitive sciences*, *4*(4), 429–452.

Di Paolo, E. A., Buhrmann, T., & Barandiaran, X. (2017). *Sensorimotor life: An enactive proposal.* Oxford: Oxford University Press.

Di Paolo, E. A., Rohde, M., & De Jaegher, H. (2010). Horizons for the enactive mind: Values, social interaction, and play. In J. Stewart, O. Gapenne, & E. A. Di Paolo (Eds.), *Enaction: Towards a new paradigm for cognitive science.* Cambridge, MA: MIT Press.

Dotov, D. (2014). Putting reins on the brain. How the body and environment use it. *Frontiers in human neuroscience*, *8*.

Dotov, D., & Chemero, A. (2014). Breaking the perception-action cycle: Experimental phenomenology of non-sense and its implications for theories of perception and movement science. In M. Cappuccio & T. Froese (Eds.), *Enactive cognition at the edge of sense-making* (pp. 37–60). Springer.

Downey, A. T. (2017). Radical sensorimotor enactivism & predictive processing. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing* (chap. 11). Frankfurt am Main: MIND Group. doi: 10.15502/9783958573123

Dretske, F. I. (1980). The intentionality of cognitive states. *Midwest studies in philosophy*, *5*(1), 281–294.

Dreyfus, H. (1991). *Being-in-the-world: A commentary on Heidegger's Being and Time, Division I.* MIT Press.

Dreyfus, H. (1992/1972). *What computers still can't do: A critique of artificial reason.* Cambridge, MA: MIT Press.

Dreyfus, H. (2002). Intelligence without representation – Merleau-Ponty's critique of mental representation: The relevance of phenomenology to scientific explanation. *Phenomenology and the Cognitive Sciences*, *1*(4), 367–383.

Dreyfus, H. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Philosophical psychology*, *20*(2), 247–268.

Dreyfus, H., & Kelly, S. D. (2007). Heterophenomenology: Heavy-handed sleight-of-hand. *Phenomenology and the Cognitive Sciences*, *6*(1-2), 45–55.

Dreyfus, H., & Taylor, C. (2015). *Retrieving realism.* Cambridge, MA: Harvard University Press.

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

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, *2*(10), 704-716.

Engel, A. K., Friston, K., Kelso, J., König, P., Kovacs, I., MacDonald III, A., … others (2010). Coordination in behavior and cognition. In C. von der Malsburg, W. A. Phillips, & W. Singer (Eds.), *Dynamic coordination in the brain.* Cambridge, MA: MIT Press.

Eyal, N. (2014). *Hooked: How to build habit-forming products.* London: Penguin.

Figee, M., Luigjes, J., Smolders, R., Valencia-Alfonso, C.-E., van Wingen, G., de Kwaasteniet, B., … others (2013). Deep brain stimulation restores frontostriatal network activity in obsessive-compulsive disorder. *Nature neuroscience*, *16*(4), 386–387.

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.

Freeman, W. J. (1987). Simulation of chaotic EEG patterns with a dynamic model of the olfactory system. *Biological cybernetics*, *56*(2-3), 139–150.

Freeman, W. J. (2000). *How brains make up their minds.* New York: Columbia University Press.

Frijda, N. H. (1986). *The emotions: Studies in emotion and social interaction.* Cambridge: Cambridge University Press.

Frijda, N. H. (2007). *The laws of emotion.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Friston, K. J. (1997). Transients, metastability, and neuronal dynamics. *Neuroimage*, *5*(2), 164–171.

Friston, K. J. (2000a). The labile brain. II. Transients, complexity and selection. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *355*(1394), 237–252.

Friston, K. J. (2000b). The labile brain. I. Neuronal transients and nonlinear coupling. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *355*(1394), 215–236.

Friston, K. J. (2008). Hierarchical models in the brain. *PLoS Comput Biol*, *4*(11), e1000211.

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

Friston, K. J. (2011). Embodied inference: Or I think therefore I am, iff I am what I think. In W. Tschacher & C. Bergomi (Eds.), *The implications of embodiment (cognition and communication)* (pp. 89–125). Imprint Academic.

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

Friston, K. J. (2012b). Policies and priors. In B. Gutkin & S. H. Ahmed (Eds.), *Computational neuroscience of drug addiction* (pp. 237–283). New York, NY: Springer.

Friston, K. J. (2013a). Active inference and free energy. *Behavioral and Brain Sciences*, *36*(03), 212–213.

Friston, K. J. (2013b). Life as we know it. *Journal of the Royal Society Interface*, *10*(86), 20130475.

Friston, K. J. (2017). Sentient dynamics: the kiss of chaos* through the Markov blanket. *Chaos and Complexity Letters*, *11*(1), 117–140.

Friston, K. J., Adams, R., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: saccades as experiments. *Frontiers in psychology*, *3*, 151.

Friston, K. J., Breakspear, M., & Deco, G. (2012). Perception and self-organized instability. *Frontiers in computational neuroscience*, *6*, 44.

Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference? *PloS one*, *4*(7), e6421.

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biological cybernetics*, *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), 1–49.

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

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

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). Retrieved from `http://rsif.royalsocietypublishing.org/content/12/105/20141383` doi: 10.1098/rsif.2014.1383

Friston, K. J., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, *104*(1), 137–160. doi: 10.1007/s00422-011-0424-z

Friston, K. J., 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., Rosch, R., Parr, T., Price, C., & Bowman, H. (2017). Deep temporal models and active inference. *Neuroscience & Biobehavioral Reviews*.

Friston, K. J., Samothrakis, S., & Montague, R. (2012). Active inference and agency: Optimal control without cost functions. *Biological cybernetics*, *106*(8-9), 523–541.

Friston, K. J., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2013). The anatomy of choice: active inference and agency. *Frontiers in Human Neuroscience*, *7*(598).

Friston, K. J., Sengupta, B., & Auletta, G. (2014). Cognitive dynamics: From attractors to active inference. *Proceedings of the IEEE*, *102*(4), 427–445.

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

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, *159*(3), 417–458.

Friston, K. J., Tononi, G., Reeke, G., Sporns, O., & Edelman, G. M. (1994). Value-dependent selection in the brain: Simulation in a synthetic neural model. *Neuroscience*, *59*(2), 229–243.

Friston, K. J., Trujillo-Barreto, N., & Daunizeau, J. (2008). Dem: a variational treatment of dynamic systems. *Neuroimage*, *41*(3), 849–885.

Frith, C. (2013). *Making up the mind: How the brain creates our mental world.* Oxford: Blackwell Publishing.

Froese, T., & Ikegami, T. (2013). The brain is not an isolated "black box," nor is its goal to become one. *Behavioral and Brain Sciences*, *36*(3), 213–214.

Froese, T., & Stewart, J. (2010). Life after Ashby: Ultrastability and the autopoietic foundations of biological autonomy. *Cybernetics & Human Knowing*, *17*(4), 7–49.

Fulda, F. C. (2017). Natural agency: The case of bacterial cognition. *Journal of the American Philosophical Association*, 1–22.

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in cognitive sciences*, *4*(1), 14–21.

Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind.* Oxford: Oxford University Press.

Gallagher, S., & Bower, M. (2014). Making enactivism even more embodied. *AVANT. Pismo Awangardy Filozoficzno-Naukowej*(2), 232–247.

Gallagher, S., & Zahavi, D. (2007). *The phenomenological mind: An introduction to philosophy of mind and cognitive science.* London: Routledge.

Gallese, V., & Sinigaglia, C. (2010). The bodily self as power for action. *Neuropsychologia*, *48*(3), 746–755.

Gallistel, C. R. (1980). *The organization of action: A new synthesis.* Hillsdale, NJ: Erlbaum.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston, MA: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston, MA: Houghton Mifflin.

Godfrey-Smith, P. (2001). On the status and explanatory structure of developmental systems theory. *Cycles of contingency: developmental systems and evolution*, 283–297.

Golonka, S. (2015). Laws and conventions in language-related behaviors. *Ecological Psychology*, *27*(3), 236–250.

Golonka, S., & Wilson, A. D. (2016). Ecological representations. *bioRxiv*, 058925.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *290*(1038), 181–197.

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *352*(1358), 1121–1127.

Greicius, M. D., Supekar, K., Menon, V., & Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral cortex*, *19*(1), 72–78.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and brain sciences*, *27*(03), 377–396.

Haken, H. (1983). *Synergetics, an introduction: Nonequilibrium phase transitions and self-organization in physics, chemistry, and biology.* New York, NY: Academic Press.

Haken, H. (2004). *Synergetics: Introduction and advanced topics.* Berlin-Heidelberg: Springer.

Haken, H., Kelso, J. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological cybernetics*, *51*(5), 347–356.

Haken, H., & Tschacher, W. (2010). Theoretical model of intentionality. *Mind and Matter*, *8*(1), 7–18.

Hatfield, G. C. (2009). *Perception and cognition: Essays in the philosophy of psychology.* Oxford: Clarendon Press.

Heft, H. (2001). *Ecological psychology in context: James gibson, roger barker, and the legacy of william james's radical empiricism.* Mahwah, NJ: Lawrence Erlbaum Associates.

Heidegger, M. (1996/1927). *Being and Time.* Oxford: Blackwell Publishing. (Macquarrie, J. & Robinson, E., trans)

Helmholtz, H. (1867). *Handbuch der physiologischen optik.* Leipzig: Leopold Voss.

Hohwy, J. (2007). The sense of self in the phenomenology of agency and perception. *Psyche*, *13*(1), 1–20.

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

Hohwy, J. (2014). The self-evidencing brain. *Noûs*, *50*(2), 259-285.

Hohwy, J. (2017). How to entrain your evil demon. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing* (chap. 2). Frankfurt am Main: MIND Group. doi: 10.15502/9783958573048

Hohwy, J., Paton, B., & Palmer, C. (2015). Distrusting the present. *Phenomenology and the Cognitive Sciences*, *15*(3), 315–335.

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 and Medicine*, *5*(6), 60–73.

Hristovski, R., Davids, K. W., & Araujo, D. (2009). Information for regulating action in sport: Metastability and emergence of tactical solutions under ecological constraints. *Perspectives on cognition and action in sport*, 43–57.

Hurley, S. L. (1998). *Consciousness in action.* Cambridge, MA: Harvard University Press.

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

Hutto, D. D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content.* MIT Press.

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

Huygens, C. (1673). *Horologium oscillatorium.* Parisiis: Apud F. Muquet. (English translation: The Pendulum Clock (Iowa State University Press, Ames, 1986).)

Iberall, A. (1977). A field and circuit thermodynamics for integrative physiology. I. Introduction to the general notions. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *233*(5), R171–R180.

Ingold, T. (2000). *The perception of the environment: essays on livelihood, dwelling and skill.* London, UK: Routledge; Psychology Press.

Ingold, T. (2011). *Being alive: Essays on movement, knowledge and description.* Taylor & Francis.

Ingold, T. (2013). *Making: Anthropology, archaeology, art and architecture.* Routledge.

Ito, J., Roy, S., Liu, Y., Cao, Y., Fletcher, M., Lu, L., … Heck, D. (2014). Whisker barrel cortex delta oscillations and gamma power in the awake mouse are linked to respiration. *Nature communications*, *5*(3572).

Jonas, H. (1966/2001). *The phenomenon of life: Toward a philosophical biology.* Evanston, IL: Northwestern University Press.

Käufer, S., & Chemero, A. (2015). *Phenomenology: an introduction.* Hoboken, NJ: John Wiley & Sons.

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

Kelso, J. S. (2012). Multistability and metastability: Understanding dynamic coordination in the brain. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *367*(1591), 906–918.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS computational biology*, *4*(11), e1000209.

Kiebel, S. J., Von Kriegstein, K., Daunizeau, J., & Friston, K. J. (2009). Recognizing sequences of sequences. *PLoS Comput Biol*, *5*(8), e1000464.

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

Kirchhoff, M. D. (2016). Autopoiesis, free energy, and the life–mind continuity thesis. *Synthese*, 1–22.

Kiverstein, J. (2006). *Naturalising phenomenology: using phenomenology to close the explanatory gap* (Unpublished doctoral dissertation). University of Edinburgh.

Kiverstein, J., & Rietveld, E. (2012). 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.* London: Palgrave Macmillan.

Kripke, S. A. (1982). *Wittgenstein on rules and private language: An elementary exposition.* Cambridge, MA: Harvard University Press.

Kruyswijk, M. (2017, Jul). Amsterdam onderhandelt met google over 'love lock bridge'. *Parool.* Retrieved from `https://www.parool.nl/amsterdam/amsterdam-onderhandelt-met-google-over-love-lock-bridge a4504901/`

Kugler, P. N., & Turvey, M. T. (1988). Self-organization, flow fields, and information. *Human Movement Science*, *7*(2-4), 97–129.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Laland, K., Matthews, B., & Feldman, M. W. (2016). An introduction to niche construction theory. *Evolutionary ecology*, *30*(2), 191–202.

Latash, M. L. (1996). The Bernstein problem: how does the central nervous system make its choices. *Dexterity and its development*, 277–303.

Lee, D. N., & Reddish, P. E. (1981). Plummeting gannets: a paradigm of ecological optics. *Nature*, *293*(5830), 293–294.

Limanowski, J., & Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Frontiers in Human Neuroscience*, *7*, 547. doi: 10.3389/fnhum.2013.00547

Marreiros, A. C., Kiebel, S. J., Daunizeau, J., Harrison, L. M., & Friston, K. J. (2009). Population dynamics under the Laplace assumption. *Neuroimage*, *44*(3), 701–714.

Maslow, A. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370-96.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living.* Berlin: Springer Science & Business Media.

Maturana, H. R., & Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding.* Boston: New Science Library/Shambhala Publications.

McDowell, J. (1994). *Mind and world.* Boston, MA: Harvard University Press.

McGregor, S., Baltieri, M., & Buckley, C. L. (2015). A minimal active inference agent. *arXiv preprint arXiv:1503.04187*.

Merleau-Ponty, M. (1942/1966). *The structure of behavior.* Boston: Beacon Press. (Fisher, Alden L., trans.)

Merleau-Ponty, M. (1945/1962). *Phenomenology of perception.* London, UK: Routledge. (Smith, C., trans)

Merleau-Ponty, M. (2003). *Nature: Course notes from the college de france.* Evanston, IL:: Northwestern University Press. (Vallier, R., trans)

Metzinger, T. (2009). *The ego tunnel: The science of the mind and the myth of the self.* New York: Basic Books.

Metzinger, T., & Wiese, W. (2017). *Philosophy and predictive processing.* Frankfurt am Main: MIND Group.

Mol, A. (2002). *The body multiple: Ontology in medical practice.* Durham, NC: Duke University Press.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective.* Cambridge, MA: MIT press.

Newsome, W. T., & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, *8*(6), 2201–2211.

Noe, A. (2004). *Action in perception.* Cambridge, MA: MIT press.

Noe, A. (2012). *Varieties of presence.* Cambridge, MA: Harvard University Press.

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

Opper, M., & Saad, D. (2001). *Advanced mean field methods: Theory and practice.* Cambridge, MA: MIT press.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, *24*(05), 939–973.

Orlandi, N. (2016). Bayesian perception is ecological perception. *Philosophical Topics*, *44*(2), 327–351.

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization studies*, *28*(9), 1435–1448.

Orr, H. A. (2009). Fitness and its role in evolutionary genetics. *Nature Reviews Genetics*, *10*(8), 531–539.

Oyama, S. (1985/2000). *The ontogeny of information: Developmental systems and evolution.* New York: Cambridge University Press.

Oyama, S. (2000). *Evolution's eye: A systems view of the biology-culture divide.* London: Duke University Press.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, *107*(1), 179–217.

Parr, T., & Friston, K. J. (2017). Working memory, attention, and salience in active inference. *Scientific Reports*, *7*(1), 14678.

Partridge, L., & Barton, N. H. (2000). Natural selection: evolving evolvability. *Nature*, *407*(6803), 457–458.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible reasoning.* Burlington, MA: Morgan Kaufmann Publishers.

Petitot, J. (1999). *Naturalizing phenomenology: Issues in contemporary phenomenology and cognitive science.* Stanford University Press.

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17–35.

Port, R. F., & Van Gelder, T. (1995). *Mind as motion: Explorations in the dynamics of cognition.* Cambridge, MA: MIT press.

Putnam, H. (1962). Brains and behavior. In R. Butler (Ed.), *Analytical philosophy: Second series* (pp. 211–235). Blackwell.

Pyragas, K. (1997). Conditional Lyapunov exponents from time series. *Physical Review E*, *56*(5), 5183.

Rabinovich, M. I., Friston, K. J., & Varona, P. (2012). *Principles of brain dynamics: global state interactions.* Cambridge, MA: MIT Press.

Rabinovich, M. I., Huerta, R., Varona, P., & Afraimovich, V. S. (2008). Transient cognitive dynamics, metastability, and decision making. *PLoS Comput Biol*, *4*(5), e1000072.

Raja, V. (2017). A theory of resonance: Towards an ecological cognitive architecture. *MInds and Machines*, 1–23.

Ramstead, M. J., Veissière, S. P., & Kirmayer, L. J. (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.

Reed, E. S. (1996). *Encountering the world: Toward an ecological psychology.* Oxford University Press.

Ridderinkhof, R. K., Forstmann, B. U., Wylie, S. A., Burle, B., & van den Wildenberg, W. P. (2011). Neurocognitive mechanisms of action control: resisting the call of the sirens. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(2), 174–192.

Rietveld, E. (2008a). Situated normativity: The normative aspect of embodied cognition in unreflective action. *Mind*, *117*(468), 973–1001.

Rietveld, E. (2008b). Special section: The skillful body as a concernful system of possible actions phenomena and neurodynamics. *Theory & Psychology*, *18*(3), 341–363.

Rietveld, E. (2008c). *Unreflective action: A philosophical contribution to integrative neuroscience* (Unpublished doctoral dissertation). ILLC-Dissertation Series DS-2008-05, University of Amsterdam.

Rietveld, E. (2012a). 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). John Benjamins Publishing Company Amsterdam/Philadelphia, PA.

Rietveld, E. (2012b). Context-switching and responsiveness to real relevance. In J. Kiverstein & M. Wheeler (Eds.), *Heidegger and cognitive science: New directions in cognitive science and philosophy* (pp. 105–135). Basingstoke, Hampshire: Palgrave Macmillan.

Rietveld, E. (2013). *Skilled intentionality for "higher" cognition.* (Unpublished Lecture at Conference, The Reach of REC (Antwerp))

Rietveld, E. (2016). Situating the embodied mind in a landscape of standing affordances for living without chairs: materializing a philosophical worldview. *Sports Medicine*, *46*(7), 927–932.

Rietveld, E., & Brouwers, A. A. (2017). Optimal grip on affordances in architectural design practices: an ethnography. *Phenomenology and the Cognitive Sciences*, *16*(3), 545–564.

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*(04), 436–436.

Rietveld, E., Denys, D., & Van Westen, M. (in press). Ecological-enactive cognition as engaging with a field of relevant affordances: The skilled intentionality framework (SIF). In A. Newen, L. de Bruin, & S. Gallagher (Eds.), *Oxford handbook of 4e cognition.* Oxford University Press.

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

Rietveld, E., Rietveld, R., & Mackic, A. (2015). The end of sitting: Towards a landscape of standing affordances. *Harv Des Mag*, *40*, 180–1.

Roesch, E. B., Nasuto, S. J., & Bishop, J. M. (2012). Emotion and anticipation in an enactive framework for cognition (response to Andy Clark). *Frontiers in psychology*, *3*.

Runeson, S. (1988). The distorted room illusion, equivalent configurations, and the specificity of static optic arrays. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(2), 295.

Runeson, S. (1989). A note on the utility of ecologically incomplete invariants. *Newsletter of the international society for ecological psychology*, *4*(1), 6–8.

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of experimental psychology: general*, *112*(4), 585.

Schneider, E. D., & Kay, J. J. (1994). Life as a manifestation of the second law of thermodynamics. *Mathematical and computer modelling*, *19*(6-8), 25–48.

Schrödinger, E. (1944). *What is life? with mind and matter and autobiographical sketches.* Cambridge, MA: Cambridge University Press.

Schwartenbeck, P., FitzGerald, T., Dolan, R. J., & Friston, K. (2013). Exploration, novelty, surprise, and free energy minimization. *Frontiers in psychology*, *4*.

Sedgwick, H. (1973). The visible horizon. *Unpublished Doctoral Dissertation, Cornell University*, *238*.

Seifert, L., Wattebled, L., Herault, R., Poizat, G., Adé, D., Gal-Petitfaux, N., & Davids, K. (2014). Neurobiological degeneracy and affordance perception support functional intra-individual variability of inter-limb coordination during ice climbing. *PloS one*, *9*(2), e89865.

Seth, A. K. (2013). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive neuroscience*, *5*(2), 97–118.

Seth, A. K. (2014). The cybernetic bayesian brain. In *Open mind.* Open MIND. Frankfurt am Main: MIND Group.

Seth, A. K. (2015). Inference to the best prediction. In *Open mind.* Open MIND. Frankfurt am Main: MIND Group.

Seth, A. K., Suzuki, K., & Critchley, H. D. (2011). An interoceptive predictive coding model of conscious presence. *Frontiers in psychology*, *2*.

Shanahan, M. (2016). The frame problem. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Spring 2016 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/spr2016/entries/frame-problem/`.

Silberstein, M. (2015). Extending neutral monism to the hard problem. *Journal of Consciousness Studies*, *22*(3-4), 181–194.

Silberstein, M., & Chemero, A. (2011). Dynamics, agency and intentional action. *Humana Mente: Journal of Philosophical Studies. Special Issue on Agency: From Embodied Cognition to Free Will. Issue*, *15*, 1–20.

Silberstein, M., & Chemero, A. (2012). Complexity and extended phenomenological-cognitive systems. *Topics in cognitive science*, *4*(1), 35–50.

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.

Stepp, N., Chemero, A., & Turvey, M. T. (2011). Philosophy for the rest of cognitive science. *Topics in Cognitive Science*, *3*(2), 425–437.

Stepp, N., & Turvey, M. T. (2010). On strong anticipation. *Cognitive Systems Research*, *11*(2), 148–164.

Stewart, J., Gapenne, O., & Di Paolo, E. A. (Eds.). (2010). *Enaction: Toward a new paradigm for cognitive science.* Cambridge, MA: MIT Press.

Stoffregen, T. A. (2003). Affordances as properties of the animal-environment system. *Ecological psychology*, *15*(2), 115–134.

Strogatz, S. H. (1994). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering.* New York, NY: Perseus Books.

Teques, P., Araújo, D., Seifert, L., del Campo, V. L., & Davids, K. (2017). The resonant system: Linking brain–body–environment in sport performance. In *Progress in brain research* (Vol. 234, pp. 33–52). Elsevier.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind.* Harvard University Press.

Thompson, E. (2011). *Mind in life and life in mind.* (Unpublished Lecture at Conference, 'The Future of the Embodied Mind' (San Sebastian))

Tognoli, E., & Kelso, J. S. (2014). The metastable brain. *Neuron*, *81*(1), 35–48.

Tschacher, W. (2010). *Glossar von fachbegriffen der theorie dynamischer systeme* (Tech. Rep.). Bern: Forschungsberichte der Universitären Pychiatrischen Dienste Bern.

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.

Tsuda, I. (2001). Toward an interpretation of dynamic neural activity in terms of chaotic dynamical systems. *Behavioral and Brain Sciences*, *24*(05), 793–810.

Turvey, M. T. (1992). Affordances and prospective control: An outline of the ontology. *Ecological psychology*, *4*(3), 173–187.

Turvey, M. T., & Carello, C. (2012). On intelligence from first principles: Guidelines for inquiry into the hypothesis of physical intelligence (PI). *Ecological Psychology*, *24*(1), 3–32.

Turvey, M. T., Shaw, R. E., Reed, E. S., & Mace, W. M. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). *Cognition*, *9*(3), 237–304.

van Dijk, L., & Rietveld, E. (2017). Foregrounding sociomaterial practice in our understanding of affordances: The skilled intentionality framework. *Frontiers in psychology*, *7*, 1969.

van Dijk, L., & Withagen, R. (2016). Temporalizing agency: Moving beyond on-and offline cognition. *Theory & Psychology*, *26*(1), 5–26.

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. (1995). What might cognition be, if not computation? *The Journal of Philosophy*, *92*(7), 345–381.

van Maanen, L., Brown, S. D., Eichele, T., Wagenmakers, E.-J., Ho, T., Serences, J., & Forstmann, B. U. (2011). Neural correlates of trial-to-trial fluctuations in response caution. *Journal of Neuroscience*, *31*(48), 17488–17495.

Van Orden, G., Hollis, G., & Wallot, S. (2012). The blue-collar brain. *Frontiers in physiology*, *3*.

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

Varela, F., Lachaux, J.-P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, *2*(4), 229-239.

Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience.* Cambridge, Massachusetts, London, England: The MIT Press.

Varga, S., & Heck, D. H. (2017). Rhythms of the body, rhythms of the brain: Respiration, neural oscillations, and embodied cognition. *Consciousness and Cognition*, *56*, 77–90.

von der Malsburg, C., Phillips, W. A., & Singer, W. (2010). *Dynamic coordination in the brain: from neurons to mind.* Cambridge, MA: MIT Press.

Von Uexküll, J. (1934/1992). A stroll through the worlds of animals and men: A picture book of invisible worlds. *Semiotica*, *89*(4), 319–391.

Walton, A. E., Richardson, M. J., Langland-Hassan, P., & Chemero, A. (2015). Improvisation and the self-organization of multiple musical bodies. *Frontiers in psychology*, *6*.

Weber, A., & Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the cognitive sciences*, *1*(2), 97–125.

Wheeler, M. (2008). Cognition in context: Phenomenology, situated robotics and the frame problem. *International Journal of Philosophical Studies*, *16*(3), 323–349.

Wiese, W., & Metzinger, T. (2017). Vanilla PP for philosophers: A primer on predictive processing. In *Philosophy and predictive processing.* MIND Group, Frankfurt am Main.

Withagen, R. (2013). *Een bewogen fundament.* 2010 Uitgevers.

Withagen, R., Araújo, D., & de Poel, H. J. (2017). Inviting affordances and agency. *New Ideas in Psychology*, *45*, 11–18.

Withagen, R., & Caljouw, S. R. (2016). 'The End of Sitting': An empirical study on working in an office of the future. *Sports Medicine*, *46*(7), 1019–1027.

Withagen, R., de Poel, H. J., Araújo, D., & Pepping, G.-J. (2012). Affordances can invite behavior: Reconsidering the relationship between affordances and agency. *New Ideas in Psychology*, *30*(2), 250–258.

Withagen, R., & Van der Kamp, J. (2010). Towards a new ecological conception of perceptual information: Lessons from a developmental systems perspective. *Human Movement Science*, *29*(1), 149–163.

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

Wittgenstein, L. (1993). *Philosophical occasions, 1912-1951.* Hackett Publishing.

Wittreich, W. J. (1952). The Honi phenomenon: a case of selective perceptual distortion. *The Journal of Abnormal and Social Psychology*, *47*(3), 705.

Wu, T. (2017). *The attention merchants: The epic scramble to get inside our heads.* London: Atlantic Books.

Yates, F. E. (2008). Homeokinetics/homeodynamics: A physical heuristic for life and complexity. *Ecological Psychology*, *20*(2), 148–179.

Zahavi, D. (2017, May). *Phenomenology of consciousness.* (Lecture at ISSA Summer school, slides retrieved from https://groups.oist.jp/issa/program)

Samenvatting

Anticipating Affordances: intentionality in self-organizing brain-body-environment systems

Het onderwerp van deze dissertatie is het wetenschappellijk en filosofisch begrijpen van *vaardige intentionaliteit.* Hiervoor worden de termen *affordance* (*handelingsmogelijkheid*) en *anticipatie* (*verwachting*) gebruikt. Vaardige intentionaliteit is het soort van intentionaliteit dat we vertonen wanneer we vaardig handelen in alledaagse situaties, zoals bijvoorbeeld wanneer we op de fiets zitten. Zulke intentionaliteit kan worden gekarakteriseerd als selectieve openheid naar affordances. Een belangrijke uitdaging voor het handelend individu is het adequaat open zijn naar affordances en het veranderen van deze openheid wanneer interacties met de omgeving minder goed gaan dan verwacht. De vraag is hoe dit mogelijk is.

Het doel van dit proefschrift is in feite tweeledig. Aan de ene kant probeert het een conceptueel kader te ontwikkelen voor het bestuderen van vaardige intentionaliteit in termen van dynamische systeem theorie en *predictive processing.* Aan de andere kant wordt in dit proefschrift beargumenteerd dat *predictive processing*, en in het bijzonder haar systeem-theoretische variant —het *free-energy principle*— het best gepaard kan gaan met een ecologische en enactivistische cognitiefilosofie in plaats van de rationalistische cognitiefilosofie waar het normaliter mee wordt geassocieerd.

Het *free-energy principle* probeert een formele en principile karakterisering te geven van de mate van wederzijdse aanpassing tussen een belichaamd individu en haar omgeving (oftewel niche). Het probeert waarnemen, handelen, leren, ontwikkeling en het veranderen van de omgeving te begrijpen als het continu verbeteren van de wederzijdse aanpassing tussen individu en omgeving. *Free-energy* is een informatie-theoretische maat voor de onderlinge afwijking tussen een individu (gekarakteriseerd als een generatief model) en een omgeving (gekarakteriseerd als een generatief proces). Filosofisch gezien deelt het *free-energy principle* zowel uitgangspunten met ecologische en enactivistische cognitiefilosofie (met name de nadruk op zelforganisatie en een adequate koppeling tussen organisme en omgeving), als met cognitivistische benaderingen (met name de nadruk op inferentie en representatie). De spanning tussen deze (ogenschijnlijk incompatibele) benaderingen is een terugkerend thema in dit proefschrift.

Dit proefschrift bevat een uitgebreide introductie waarin de centrale probleemstelling uiteen wordt gezet, gevolgd door vijf los van elkaar leesbare hoofdstukken.

In *Hoofdstuk 1* wordt een conceptueel raamwerk voor *Radical Embodied Cognitive Neuroscience* ontwikkeld. Het eerste deel van dit hoofdstuk introduceert vaardige intentionaliteit en de structuur van de ecologische niche, dat wil zeggen het 'landschap van affordances?. Het tweede deel van dit hoofdstuk verbindt vaardige intentionaliteit met theorien over zelforganisatie en neurodynamica, zoals het *free-energy principle*. Het derde deel van het hoofdstuk illustreert deze integratieve benadering door middel van het presenteren van onderzoek uit de bewegingswetenschap en de impact van diepe hersenstimulatie op de openheid naar affordances.

In *Hoofdstuk 2* wordt de relatie tussen ecologische en enactivistische cognitiefilosofie en het *free-energy principle* verder uitgewerkt. Het eerste deel van het hoofdstuk presenteert het *free-energy principle* en haar relatie tot de zogeheten *mind-life continuity thesis*. In het tweede deel van het hoofdstuk wordt de metafoor van de *crooked scientist* of *frauduleuze wetenschapper* gepresenteerd als een manier om zowel kennis als preferenties binnen n raamwerk te vatten. Het derde deel van het hoofdstuk bespreekt (en betwist) de relatie tussen internalisme, inferentie en Markov blankets.

Hoofdstuk 3 vergelijkt een rationalistische, een cybernetische en een ecologisch-enactivistische interpretatie van *predictive processing*. Hoe deze drie interpretaties omgaan met de *sense of agency* en intentionaliteit is onderwerp voor het tweede deel van dit hoofdstuk.

De structuur van de sociaal-materiele omgeving is het onderwerp van *Hoofdstuk 4*. In het hoofdstuk wordt een nieuwe theorie van ecologische informatie ontwikkeld die het mogelijk maakt om gevallen van zogeheten 'hogere' cognitie te begrijpen in termen van regelmatigheden in de menselijke levensvorm of ecologische niche.

Hoofdstuk 5 poogt door middel van computationele simulaties te laten zien dat free-energy een relationele kwantiteit is die de wederzijdse aanpassing tussen een belichaamd individu en een eco-niche beschrijft. De wiskundige overeenkomst tussen de wijze waarop een individu leert over de omgeving en de wijze waarop de omgeving 'leert' over het individu duidt op een diepe symmetrie tussen organisme en omgeving in de totstandkoming van handelingen. Of en hoe de constructie van een niche plaatsvindt hangt af van het relatieve gemak waarmee organisme en omgeving zich aanpassen als een functie van de interacties die ze hebben.

Summary

Anticipating Affordances: intentionality in self-organizing brain-body-environment systems

The topic of this dissertation is the naturalization of *skilled intentionality* in terms of the concepts of *affordance* and *anticipation.* Skilled intentionality is the kind of intentionality exhibited by skilled agents when acting in everyday situations and is characterized as selective openness to affordances. A main task of a skilled agent is to be sensitive to how well its faring in its interactions with the environment and to selectively change its openness to affordances when things are not going as well as anticipated. The question is how this is possible.

The aim of this dissertation is twofold. On the one hand, it is an attempt to develop a conceptual framework for the study of *skilled intentionality* in terms of dynamical systems theory and predictive processing. On the other hand, it presents the argument that predictive processing, and especially its systems theoretic cousin, the free-energy principle, are best understood in terms of an ecological and enactive philosophy of mind, and not in the rationalist philosophy of mind it is standardly associated with.

The free-energy principle is a principled and formal attempt to describe the 'fit' between an embodied agent and its ecological niche, and to explain how agents perceive, act, learn, develop and structure their environment in order to optimize their fitness, or minimize their free-energy. The quantity of free-energy is an information-theoretic measure that captures the discrepancy between an agent (generative model) and its niche (generative process). The free-energy principle shares with ecological and enactive philosophy of mind an emphasis on self-organization and adequate coupling with the environment, while at the same time it seems to share with more cognitivist approaches to philosophy of mind a commitment to inference and representation. I will attempt to solve this puzzle.

The dissertation contains an extensive introduction in which the stage is set, followed by five independently readable chapters.

Chapter 1 develops a conceptual framework for Radical Embodied Cognitive Neuroscience. The first part of the chapter introduces skilled intentionality and the structure of the landscape of affordances. The second part of the chapter relates skilled intentionality to theories of self-organization and neurodynamics, such as the free-energy principle. The third part of the chapter exemplifies the integrative approach by relating our conceptual framework to a study of the impact of Deep Brain Stimulation on affordance responsiveness.

Chapter 2 develops the link between ecological-enactive cognitive science and the free-energy principle. The first part of the paper presents the free-energy principle and its commitment to the mind-life continuity thesis. In the second part of the chapter, the 'crooked-scientist' metaphor is developed as a way to make sense of the integration of value and epistemics under the free-energy principle. In the third part of the chapter, the link between internalism, inference and Markov blankets is challenged by discussing coupled clocks as an intuition pump

Chapter 3 compares a rationalist Helmholtzian, a cybernetic and an ecological-enactive interpretation of predictive processing. In the second part of the chapter, the discussion focusses on how each of these three interpretations conceives of the sense of agency and intentionality in different ways.

Chapter 4 focusses on the structure of the socio-material environment. In particular, it presents a novel account of ecological information designed to work for cases of 'higher' cognition. Introducing the notion of general ecological information, an account is given of these regularities in terms of constraints, information and the form of life or ecological niche.

Chapter 5 uses computational simulations of a free-energy minimizing agent to show that free-energy is a relational quantity, pertaining to the 'fit' between an embodied agent and its econiche. A formal similarity between the way an agent remembers its environment and the way the environment 'remembers' the behavior of an agent is exploited, and it is shown how niche construction is critically dependent on the learning rate of the agent, and the 'inertia' or malleability of the environment.

Titles in the ILLC Dissertation Series:

ILLC DS-2009-01: **Jakub Szymanik**
Quantifiers in TIME and SPACE. Computational Complexity of Generalized Quantifiers in Natural Language

ILLC DS-2009-02: **Hartmut Fitz**
Neural Syntax

ILLC DS-2009-03: **Brian Thomas Semmes**
A Game for the Borel Functions

ILLC DS-2009-04: **Sara L. Uckelman**
Modalities in Medieval Logic

ILLC DS-2009-05: **Andreas Witzel**
Knowledge and Games: Theory and Implementation

ILLC DS-2009-06: **Chantal Bax**
Subjectivity after Wittgenstein. Wittgenstein's embodied and embedded subject and the debate about the death of man.

ILLC DS-2009-07: **Kata Balogh**
Theme with Variations. A Context-based Analysis of Focus

ILLC DS-2009-08: **Tomohiro Hoshi**
Epistemic Dynamics and Protocol Information

ILLC DS-2009-09: **Olivia Ladinig**
Temporal expectations and their violations

ILLC DS-2009-10: **Tikitu de Jager**
"Now that you mention it, I wonder...": Awareness, Attention, Assumption

ILLC DS-2009-11: **Michael Franke**
Signal to Act: Game Theory in Pragmatics

ILLC DS-2009-12: **Joel Uckelman**
More Than the Sum of Its Parts: Compact Preference Representation Over Combinatorial Domains

ILLC DS-2009-13: **Stefan Bold**
Cardinals as Ultrapowers. A Canonical Measure Analysis under the Axiom of Determinacy.

ILLC DS-2010-01: **Reut Tsarfaty**
Relational-Realizational Parsing

ILLC DS-2010-02: **Jonathan Zvesper**
Playing with Information

ILLC DS-2010-03: **Cdric Dgremont**
The Temporal Mind. Observations on the logic of belief change in interactive systems

ILLC DS-2010-04: **Daisuke Ikegami**
Games in Set Theory and Logic

ILLC DS-2010-05: **Jarmo Kontinen**
Coherence and Complexity in Fragments of Dependence Logic

ILLC DS-2010-06: **Yanjing Wang**
Epistemic Modelling and Protocol Dynamics

ILLC DS-2010-07: **Marc Staudacher**
Use theories of meaning between conventions and social norms

ILLC DS-2010-08: **Amlie Gheerbrant**
Fixed-Point Logics on Trees

ILLC DS-2010-09: **Galle Fontaine**
Modal Fixpoint Logic: Some Model Theoretic Questions

ILLC DS-2010-10: **Jacob Vosmaer**
Logic, Algebra and Topology. Investigations into canonical extensions, duality theory and point-free topology.

ILLC DS-2010-11: **Nina Gierasimczuk**
Knowing One's Limits. Logical Analysis of Inductive Inference

ILLC DS-2010-12: **Martin Mose Bentzen**
Stit, Iit, and Deontic Logic for Action Types

ILLC DS-2011-01: **Wouter M. Koolen**
Combining Strategies Efficiently: High-Quality Decisions from Conflicting Advice

ILLC DS-2011-02: **Fernando Raymundo Velazquez-Quesada**
Small steps in dynamics of information

ILLC DS-2011-03: **Marijn Koolen**
The Meaning of Structure: the Value of Link Evidence for Information Retrieval

ILLC DS-2011-04: **Junte Zhang**
System Evaluation of Archival Description and Access

ILLC DS-2011-05: **Lauri Keskinen**
Characterizing All Models in Infinite Cardinalities

ILLC DS-2011-06: **Rianne Kaptein**
Effective Focused Retrieval by Exploiting Query Context and Document Structure

ILLC DS-2011-07: **Jop Brit**
Grothendieck Inequalities, Nonlocal Games and Optimization

ILLC DS-2011-08: **Stefan Minica**
Dynamic Logic of Questions

ILLC DS-2011-09: **Raul Andres Leal**
Modalities Through the Looking Glass: A study on coalgebraic modal logic and their applications

ILLC DS-2011-10: **Lena Kurzen**
Complexity in Interaction

ILLC DS-2011-11: **Gideon Borensztajn**
The neural basis of structure in language

ILLC DS-2012-01: **Federico Sangati**
Decomposing and Regenerating Syntactic Trees

ILLC DS-2012-02: **Markos Mylonakis**
Learning the Latent Structure of Translation

ILLC DS-2012-03: **Edgar Jos Andrade Lotero**
Models of Language: Towards a practice-based account of information in natural language

ILLC DS-2012-04: **Yurii Khomskii**
Regularity Properties and Definability in the Real Number Continuum: idealized forcing, polarized partitions, Hausdorff gaps and mad families in the projective hierarchy.

ILLC DS-2012-05: **David Garca Soriano**
Query-Efficient Computation in Property Testing and Learning Theory

ILLC DS-2012-06: **Dimitris Gakis**
Contextual Metaphilosophy - The Case of Wittgenstein

ILLC DS-2012-07: **Pietro Galliani**
The Dynamics of Imperfect Information

ILLC DS-2012-08: **Umberto Grandi**
Binary Aggregation with Integrity Constraints

ILLC DS-2012-09: **Wesley Halcrow Holliday**
Knowing What Follows: Epistemic Closure and Epistemic Logic

ILLC DS-2012-10: **Jeremy Meyers**
Locations, Bodies, and Sets: A model theoretic investigation into nominalistic mereologies

ILLC DS-2012-11: **Floor Sietsma**
Logics of Communication and Knowledge

ILLC DS-2012-12: **Joris Dormans**
Engineering emergence: applied theory for game design

ILLC DS-2013-01: **Simon Pauw**
Size Matters: Grounding Quantifiers in Spatial Perception

ILLC DS-2013-02: **Virginie Fiutek**
Playing with Knowledge and Belief

ILLC DS-2013-03: **Giannicola Scarpa**
Quantum entanglement in non-local games, graph parameters and zero-error information theory

ILLC DS-2014-01: **Machiel Keestra**
Sculpting the Space of Actions. Explaining Human Action by Integrating Intentions and Mechanisms

ILLC DS-2014-02: **Thomas Icard**
The Algorithmic Mind: A Study of Inference in Action

ILLC DS-2014-03: **Harald A. Bastiaanse**
Very, Many, Small, Penguins

ILLC DS-2014-04: **Ben Rodenhuser**
A Matter of Trust: Dynamic Attitudes in Epistemic Logic

ILLC DS-2015-01: **Mara Ins Crespo**
Affecting Meaning. Subjectivity and evaluativity in gradable adjectives.

ILLC DS-2015-02: **Mathias Winther Madsen**
The Kid, the Clerk, and the Gambler - Critical Studies in Statistics and Cognitive Science

ILLC DS-2015-03: **Shengyang Zhong**
Orthogonality and Quantum Geometry: Towards a Relational Reconstruction of Quantum Theory

ILLC DS-2015-04: **Sumit Sourabh**
Correspondence and Canonicity in Non-Classical Logic

ILLC DS-2015-05: **Facundo Carreiro**
Fragments of Fixpoint Logics: Automata and Expressiveness

ILLC DS-2016-01: **Ivano A. Ciardelli**
Questions in Logic

ILLC DS-2016-02: **Zo Christoff**
Dynamic Logics of Networks: Information Flow and the Spread of Opinion

ILLC DS-2016-03: **Fleur Leonie Bouwer**
What do we need to hear a beat? The influence of attention, musical abilities, and accents on the perception of metrical rhythm

ILLC DS-2016-04: **Johannes Marti**
Interpreting Linguistic Behavior with Possible World Models

ILLC DS-2016-05: **Phong L**
Learning Vector Representations for Sentences - The Recursive Deep Learning Approach

ILLC DS-2016-06: **Gideon Maillette de Buy Wenniger**
Aligning the Foundations of Hierarchical Statistical Machine Translation

ILLC DS-2016-07: **Andreas van Cranenburgh**
Rich Statistical Parsing and Literary Language

ILLC DS-2016-08: **Florian Speelman**
Position-based Quantum Cryptography and Catalytic Computation

ILLC DS-2016-09: **Teresa Piovesan**
Quantum entanglement: insights via graph parameters and conic optimization

ILLC DS-2016-10: **Paula Henk**
Nonstandard Provability for Peano Arithmetic. A Modal Perspective

ILLC DS-2017-01: **Paolo Galeazzi**
Play Without Regret

ILLC DS-2017-02: **Riccardo Pinosio**
The Logic of Kant's Temporal Continuum

ILLC DS-2017-03: **Matthijs Westera**
Exhaustivity and intonation: a unified theory

ILLC DS-2017-04: **Giovanni Cin**
Categories for the working modal logician

ILLC DS-2017-05: **Shane Noah Steinert-Threlkeld**
Communication and Computation: New Questions About Compositionality

ILLC DS-2017-06: **Peter Hawke**
The Problem of Epistemic Relevance

ILLC DS-2017-07: **Aybke zgn**
Evidence in Epistemic Logic: A Topological Perspective

ILLC DS-2017-08: **Raquel Garrido Alhama**
Computational Modelling of Artificial Language Learning: Retention, Recognition & Recurrence

ILLC DS-2017-09: **Milo Stanojevi**
Permutation Forests for Modeling Word Order in Machine Translation

ILLC DS-2018-01: **Berit Janssen**
Retained or Lost in Transmission? Analyzing and Predicting Stability in Dutch Folk Songs

ILLC DS-2018-02: **Hugo Huurdeman**
Supporting the Complex Dynamics of the Information Seeking Process

ILLC DS-2018-03: **Corina Koolen**
Reading beyond the female: The relationship between perception of author gender and literary quality

ISBN 978-90-90-30840
www.illc.uva.nl