Health, Agency, and the Evolution of Consciousness

Walter Veit

BA in Philosophy & Economics,
University of Bayreuth, Germany, 2018

MA in Philosophy of Biological and Cognitive Sciences,
University of Bristol, UK, 2019

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

School of History and Philosophy of Science

Faculty of Science

The University of Sydney

2022

Health, Agency, and the Evolution of Consciousness

Walter Veit

Abstract

This goal of this thesis in the philosophy of nature is to move us closer towards a true biological science of consciousness in which the evolutionary origin, function, and phylogenetic diversity of consciousness are moved from the field's periphery of investigations to its very centre. Rather than applying theories of consciousness built top-down on the human case to other animals, I argue that we require an evolutionary bottom-up approach that begins with the very origins of subjective experience in order to make sense of the place of mind in nature. To achieve this goal, I introduce and defend the *pathological complexity thesis* as both a framework for the scientific investigation of consciousness and as a life-mind continuity thesis about the origins and function of consciousness.

Dedication

I dedicate this thesis to my parents Ieva and Werner.

Declaration

I declare that this thesis is my own work, written while I was a PhD student in the School of History and Philosophy of Science, The University of Sydney. To the best of my knowledge, all assistance I have received and the sources I have used have been acknowledged. The thesis has not been submitted for any other degree and I have no conflicts of interest to declare.

Parts of this thesis are published or in press. An article based on Chapter 1 will appear in Synthese (Veit forthcoming).

Chapter 3 is a longer and more extensive version of Veit (2022c). Section 2.6 also reuses some material from this article.

Chapter 4 is an expanded and improved version of Veit (2022a). Sections 1.1., 1.3, 6.3, and 6.4 also reuses some of the material from this article.

Chapter 5 is an improved and expanded version of Veit (2022f), with added discussions of octopuses, fish, reptiles and birds. Section 1.1 also uses some of the content from this article.

Finally, Section 3.2.2 also includes material from Veit (2022b) and Section 5.4 uses some of the material from Veit (2021b).

Walter Veit

Acknowledgements

This thesis this is the result of several years of thinking about evolution, agency, health, consciousness, and animal life. While philosophy is often described as an ivory tower activity, this could be no further from the truth for naturalistic philosophy, and I have benefited greatly from the advice and feedback from scientists and philosophers alike. During this time, I have been been in contact with more people than I could mention here, so I preemptively ask forgiveness from anyone I have forgotten to mention.

Beginning with my advisors, I would like to thank all of them for their help and consistently constructive criticisms. Firstly, my primary supervisor Paul Griffiths, who helped the development of this thesis immensely with his wide range of expertise, humour, and ultimately funding from his ambitious 'A Philosophy of Medicine for the 21st century' project that supported my PhD (for which I acknowledge funding from the Australian Research Council's Discovery Projects funding scheme [project number FL170100160]). Secondly, Peter Godfrey-Smith, whose work inspired me to write a thesis on animal consciousness, and whose feedback and extensive knowledge of the field have been incredibly helpful. Thirdly, I would like to thank Marian Dawkins, Professor of Animal Behaviour at the University of Oxford, who kindly agreed to serve as an external advisor for roughly the last half of the duration of my PhD. While we eventually gave up on formalizing this status due to excessive paperwork, she nevertheless deserves mention here for her role as an informal advisor and her helpful feedback on this thesis.

Furthermore, I have benefited from exchanges with other members of the University of Sydney. I might mention, amongst others, Maureen O'Malley, Dominic Murphy, and Ofer Gal. Being part of Griffiths' wonderful interdisciplinary Theory and Method in Biosciences group at the Charles Perkins Centre - one of Australia's leading medical research institutes - enabled me to present my work multiple times to an entire group of philosophers of biology and to grow into an interdisciplinary researcher with a firm understanding of evolution. Indeed, this thesis very much reflects the mission of this 'philosophy lab' to employ the integrative power of biological theory and especially evolutionary biology to remove "conceptual and methodological roadblocks to the advancement of science".[1]

Next, I would like to acknowledge four host institutions that I visited during my PhD. Firstly, I would like to thank Jonathan Birch, who hosted me for a term in his Animal Sentience team within the Centre for Philosophy of Natural and Social Science at the London School of Economics, which provided the perfect place to discuss the ideas of this thesis. Secondly, I would like to thank Stephan Hartmann, who hosted me for roughly two months at the Munich Center for Mathematical

[1] https://tmbiosci.org/ [Accessed: July 6th, 2022]

Philosophy at the Ludwig-Maximilians-Universität München, where I delved deep into the possibility of a mathematical framework for consciousness, taught a course in the history and philosophy of biology, and was eventually honoured with the status of external member. Thirdly, I would like to thank Nicola Clayton for letting me visit her Comparative Cognition Lab at the University of Cambridge, Department of Psychology for five months. Observing actual research on animal cognition and engaging with researchers at her lab has been incredibly helpful in the writing of this thesis. Lastly, I would like to thank Rob Salguero-Gómez for hosting me in his life history team at the University of Oxford, Department of Zoology in the last months before submitting my thesis.

I have also presented ideas of this thesis to audiences at meetings of the Association for the Scientific Study of Consciousness, the International Network of Economic Method, the International Society for the History, Philosophy, and Social Studies of Biology, the Australasian Association of Philosophy, the Joint Brazilian Annual Ethological/Latin American Ethological Conference, and the Power of Scents workshop Sokoine organized by APOPO at the University of Agriculture in Tanzania, which was followed by a Safari in the Mikumi National Park - the perfect place to think about animal life. Furthermore, I also would like to thank audiences at a keynote I gave at the 16th UFAW Animal Welfare Student Conference, and research seminars at the LMU and the University of Bayreuth.

For comments (regarding both content and grammar) on parts of this thesis, I owe gratitude to Daniel Dennett, Jonathan Birch, Kristin Andrews, David Spurrett, Don Ross, Colin Allen, Stefan Gawronski, Kate Lynch, Peter Takacs, Samir Okasha, Carolyn Ristau, as well as reviewers for the article versions of my chapters. Furthermore, I would like to express my thanks to anonymous reviewers at both Routledge and Palgrave Macmillan for their feedback on a book version of this thesis that helped to improve the thesis itself.

Lastly, I would like to thank my fiancée and frequent collaborator Heather Browning for her unending love, brilliant feedback, and thorough proofreading and copyediting. This thesis has greatly benefited from innumerable discussions with a philosopher of her caliber specializing in animal sentience and welfare that helped to refine many of my arguments presented in this thesis.

Contents

List of Figures

List of Tables

Introduction

> *Origins of Man now proved.— Metaphysics must flourish.— He who understands baboon would do more towards metaphysics than Locke.*
>
> – Charles Robert Darwin (1838)

In Search of the Place of Mind in Nature

The target phenomenon of this thesis is one that could hardly constitute a greater challenge to a paleobiologist. It is a phenomenon that is said to leave no fossil trace and has repeatedly been described as the hardest problem of biology: consciousness. In nature, we seem to find a striking difference between systems without any sort of conscious experience, like Australian bush fires, the plants that succumb to them, robots, bacteria, our planet, other stars, and the universe as a whole; and those systems, like humans, for which there is something it is like to be them - or so most take for granted. When we ask whether there is something it is like to be a bat, or any other living organism for that matter, we are asking both whether they have subjective experiences (of any kind) and what these experiences consist in. Yet, how could we possibly learn about the nature of this elusive phenomenon?

The problems of consciousness have for a long time puzzled both scientists and philosophers, even deemed exceedingly difficult if not impossible to answer: What is consciousness and why does it exist at all? Could consciousness come in degrees and different variations or is it like a light-switch that is either 'on' or 'off'? Finally, which animals are conscious and do they differ in their subjective experiences? Are humans the only conscious beings on our planet? Or should we include all mammals? Birds as well? Or all the animals? Why not say that all life is sentient? This view is sometimes called *biopsychism* and though it will strike many as surely too strong, that does not mean that it lacks defenders.[2]

The German biologist, philosopher, and artist Ernst Haeckel (1892) - who is sometimes described as the "German Darwin" for his devout defense and development of Darwin's theoretical framework in Germany[3] - was the one who coined the term 'biopsychism'. But he eventually went on to defend an even broader view called *panpsychism*: the view that "[a]ll matter is ensouled" and that feeling should be conceived of as "a universal world-principle" (1892, p. 483). Contemporary panpsychists think that our fundamental scientific image of physics needs to be radically updated to include an aspect or degree of mentality in all matter, such as

[2]Consider for instance Herbert S. Jennings (1904), Henri Bergson (1920), Maxine Sheets-Johnstone (1999), Lynn Margulis (2001), and Arthur Reber (2019).

[3]See Aveling (1886); Kutschera et al. (2019).

electrons, in order to make sense of the presence of minds; a view that even its proponents admit is readily rejected by most philosophers and non-philosophers alike as - to put it bluntly - absurd (Goff et al. 2020). Yet, the view has very prominent defenders such as Thomas Nagel and David Chalmers, who have been incredibly influential in the shaping of the philosophical and scientific discourse around consciousness and urge us to take this radical option seriously. They hold that the problems of panpsychism are no more serious than those for any other view, though Nagel (1986) also admits that the view has "the faintly sickening odor of something put together in the metaphysical laboratory" (p. 49). But while it appears easy enough to dismiss such radical views as going too far, it is also apparent that there is little agreement on how we could even *possibly* settle the question.

Disagreement about the possibility of pain in fish or in invertebrates such as insects sometimes appears no less contested than the metaphysical view that consciousness pervades the universe. Largely, this is due to a worry the British biologist Thomas Huxley - also known as *Darwin's bulldog* - once famously expressed:

> [W]hat consciousness is, we know not; and how it is that anything so remarkable as a state of consciousness comes about as the result of irritating nervous tissue, is just as unaccountable as the appearance of the Djin when Aladdin rubbed his lamp in the story, or as any other ultimate fact of nature.
>
> – Thomas Henry Huxley (in Huxley and Youmans 1868, p. 178)

Huxley did not perceive how we could possibly explain consciousness as a causally efficacious materialist phenomenon, which led him to endorse epiphenomenalism - i.e. the dualist view that subjective experience is only the effect, never the cause of the physical processes of the brain.[4] His concerns were later better articulated by Joseph Levine (1983), who expressed the mind-body problem in terms of what he called "the explanatory gap" between the mental and the physical - an epistemological rather than ontological gap he thought could only be bridged by eliminating the mental.[5] That would be a radical non-dualist answer to the mind-body problem that many would consider to be too 'hard-headed'; reductive materialism gone too far. How could we possibly eliminate the most cherished and directly experienced aspect of our mental lives? The framing now more commonly used in debates about consciousness is Chalmers's (1995) description of this explanatory gap as the so-called 'hard problem of consciousness'. Chalmers maintained that while we can readily make progress on the 'easy problems of consciousness', i.e. the functional, computational, and mechanistic side of the mind, using the standard tools and methods of cognitive science, none of this appears to address the hard problem of how it generates a first-person phenomenological feel of mental phenomena:

> What makes the hard problem hard and almost unique is that it goes *beyond* problems about the performance of functions. To see this, note that even when

[4] See Huxley (2011) and Campbell (2001) for a deeper discussion of Huxley's views.

[5] Godfrey-Smith (2020b) contests whether Huxley expressed an early version of the explanatory gap. Yet, the parallels are quite striking. Whereas Huxley found himself led towards epiphenomenalism, Levine suggested that the only way to close the gap was to become an eliminative materialist in regards to the qualitative side of the mind, though admitting that it has so far remained stubbornly "resistant to philosophical attempts" at elimination (Levine 1983, p. 361). They, nevertheless, both saw the mental side as a deeply problematic explanatory challenge without any intuitively attractive solutions.

> we have explained the performance of all the cognitive and behavioral functions in the vicinity of experience—perceptual discrimination, categorization, internal access, verbal report—there may still remain a further unanswered question: *Why is the performance of these functions accompanied by experience?* A simple explanation of the functions leaves this question open.
>
> – David Chalmers (1995, p. 202) [emphasis in original]

This framing of the problem of consciousness makes it appear as if science could never address the qualitative feel, i.e. the 'qualia', of subjective experience. Following Nagel's (1974) famous essay 'What Is It Like to Be a Bat?', in which he set out to argue that while bats are likely conscious we could never know what their subjective experience was like, phenomenological properties or qualia are now typically treated as something like a second-order property of what it is like or what it feels like to have a mental state, as opposed to a first-order property of the state itself (Sytsma and Machery 2010, p. 299). Thanks to Chalmers' and Nagel's influence on the philosophy of mind, it is now typical to consider the problem of consciousness as identical to the problem of qualia, rather than of a particularly rich cognitive phenomenon with qualitative aspects that may be unique to humans, as it was in the literature of the philosophy of mind in the 1980s (Godfrey-Smith 2016c,a, 2017a). Could this combination of two formerly distinct problems have given rise to the increasing conviction among many philosophers and scientists that there is something like a hard problem that cannot be solved?

An even more radical rejection of the idea that science could provide a materialist account of consciousness can be found in the form of *idealism*, i.e. the notion that everything is mental. What are we to make of this, what Godfrey-Smith (2020b) candidly called a "wild sweep of these alternative views of the universe" (p. 14)? In thinking about the place of mind in nature there almost appears to be a dreadful possibility that *anything goes* and this is not restricted to 'mere' philosophical discussions about the very nature of mind. If we ask about the presence of other minds in non-humans it appears that we are faced with just as much uncertainty as with the big-picture view about the relationship between matter and mind; at least two senses in which we can ask for the place of mind in nature.

How are we to respond to a biopsychist who points to the autonomy, sophisticated sensory feedback, and decision-making of single-celled organisms? To assert that all of the actions of an animal could be explained by mere mechanics without the presence of mind is a tactic frequently used in discussions over all kinds of possible boundaries, including fish, insects, and mammals. Why is someone wrong who denies pain in octopuses or crabs? How could we possibly settle these debates about where to draw the line? Many answers to this question of the place of mind in nature have been proposed, such as the eliminativist or illusionist view that *no one* has consciousness in the sense of possessing qualia (Levine 1983; Dennett 1991; Frankish 2017), the exclusive attribution of consciousness to humans (Macphail 1998), only to the great apes (Bermond 2001), only to mammals and birds (Edelman and Tononi 2000), to all mammals, birds, and non-avian reptiles (Cabanac et al. 2009), to all vertebrates (Mashour and Alkire 2013), to all vertebrates as well as some invertebrate groups such as cephalopods, crustaceans, and insects (Ginsburg and Jablonka 2010; Bronfman et al. 2016; Barron and Klein 2016; Tye 2016), to plants as well (Gagliano 2017; Trewavas et al. 2020), to all living organisms including single-celled

ones (Margulis 2001; Reber 2019), and to all entities in the universe (Goff et al. 2020).[6] Views on the presence of consciousness range from none to all.

We are faced with such a diversity of alternative models of consciousness that it almost seems like we have an embarrassment of riches. Without a *standard* for thinking about these problems of the mind, it almost looks like, as Godfrey-Smith (2020b) put it, "[p]eople can say whatever they like" (p. 15). This view of philosophy as a state of indefinite arbitrariness is to be strongly resisted. Following a long tradition of naturalist thinkers, this thesis firmly rejects the view common in some areas of philosophy that our profession is primarily engaged in a game of mere conceptual exploration; that philosophy is merely concerned with expanding the space of *possible* views.[7] The following chapters constitute an exercise in naturalistic philosophy to make sense of the place of consciousness in nature by providing the science of consciousness with a much-needed *standard* that it is unfortunately still lacking.

A Darwinian *Standard*

As other Darwinian thinkers have argued, this standard should not be the cherished insight derived from human first-person experience, but the modern twenty-first century theory of evolutionary biology. It is only by investigating the evolutionary *origins* of consciousness and the ecological *lifestyles* of these first conscious entities, that we will truly understand the place of consciousness in nature without being misled by the particularities, idiosyncrasies, and complexities of the human mind. The shared ancestry of all life on Earth provides us with a rich set of theoretical tools and constraints with which to understand the origins of biological phenomena. And of course, consciousness is just that, an evolved biological phenomenon - something that is now widely accepted among both philosophers and scientists writing about consciousness.

So one would be led to believe, as the neuroscientist Simona Ginsburg and evolutionary biologist Eva Jablonka note in their recent 2019 treatise *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*, that philosophers and scientists alike had firmly integrated evolutionary theory into "the framework of consciousness studies, both as a yardstick for measuring the validity of new theories and as a source of insights" (p. x). But, despite the efforts of many Darwinian thinkers, this has not happened. As the epigraph at the beginning this introduction was meant to emphasize, Darwin had already realized twenty-one years prior to the publication of *The Origin of Species*, in a little private notebook[8], that a view of life in terms of shared ancestry would radically transform our view of nature and our place within it. Yet, evolutionary and ecological thinking about the role of consciousness in nature has played a surprisingly small role in the study of consciousness. Ginsburg and Jablonka describe this lack of Darwinian thinking within a supposedly naturalistic study of consciousness highly critically: "until very recently there has been a strange lacuna in the field. Although most scientists and

[6]This list is not meant to be exhaustive. See Liljenström and Århem (2008) for a collection of essays and Griffin (2001); Ginsburg and Jablonka (2010, 2019); Dawkins (2021) for other expositions of alternative views on the phylogenetic distribution of consciousness that my list draws on.

[7]I thank Kim Sterelny for raising this point against the panpsychist literature at a bonfire at the *Philosophy of Biology at Dolphin Beach* (PBDB) workshop in 2020.

[8]See *Notebook M* (Darwin 1838).

philosophers who write about consciousness are now convinced that it is a biological process that is a product of evolution, its evolutionary origins are rarely central to their discussions" (p. x).

Undoubtedly, this can be explained through the perceived difficulty and speculative nature of adaptationist reverse-engineering approaches to the mind - especially the human mind - which have been sneered at as 'just-so stories': plausibly sounding explanations that aren't empirically testable (Gould and Lewontin 1979). While there have been innumerable attempts at a functionalist approach to consciousness, due to its status as a standard objection to epiphenomnalist views that make its evolution a mystery,[9] such thinking has unfortunately often avoided an investigation of its evolutionary origins in other animals, instead focusing on humans and humans alone.

Yet, it is no surprise that such evolutionary explanations would be avoided in a scientific investigation that was *already* seen as deeply suspect due to the lingering after-effects of the behaviourist project that banished consciousness from science. But this neglect of Darwinian thinking is unfortunate because evolutionary theory provides us with both a rich theoretical framework for thinking about consciousness and an important set of constraints that any theory of consciousness should account for. If we can build a theory of consciousness that doesn't leave its evolutionary origins a mystery, one that can explain the dawn of qualia, then we will no longer be in a position where people could say that no view is better than any other and all cards should be left on the table as equal contenders. A theory that can explain the evolution of consciousness in a gradual fashion through small incremental steps is to be preferred over any theory that demands a dualist carving of nature at its joints; a big jump or a sudden explosion of mindedness. Instead, we would end up with a historical explanation of the place of consciousness as a complex phenomenon in nature that at least substantially narrows the explanatory gap between matter and mind. As Ginsburg and Jablonka put it:

> Evolutionary theory is [...] the most general framework for understanding the biological world. It is a conceptual bottleneck through which any theory of life and mind must pass. If a biological (or psychological, or sociological) theory fails to pass through this bottleneck, it is likely there is something seriously wrong with it.
>
> – Simona Ginsburg and Eva Jablonka (2019, pp. ix-x)

Like most scientists studying consciousness, this thesis will treat consciousness as a complex evolved biological phenomenon related to the brain and nervous system of animals; something that was *built* over aeons of evolutionary time. But if consciousness is a biological phenomenon, then it ought to be treated as such. Whereas some prominent figures such as Nagel (2012) see the problems of consciousness as a fundamental flaw in evolutionary theory, their views have shown a striking lack of knowledge and under-utilization of the theoretical toolkit modern evolutionary biology has to offer.

If we are interested in the place of mind in nature, we must place the question of its origin, function, and phylogenetic diversity across the tree of life at the very heart

[9] See the target article of Dawkins (1990) and the commentaries to it for an excellent discussion of epiphenomenalism and the causal role of subjective feelings.

of a true biological science of consciousness. It is only by asking the functionalist question of what consciousness in all of its varieties and gradations does *for* healthy sentient agents within their normal ecological lifestyles and the natural environments they have evolved in, that we can transition towards a true biological study of consciousness. This naturalist endeavour is fundamentally what *Health, Agency, and the Evolution of Consciousness* will attempt to accomplish.

While there have been numerous attempts to address the problems of consciousness through a functionalist/evolutionary approach, (many of which I will draw on throughout this thesis), my approach will stand out by offering a new strategy at making progress on these problems through an emphasis of animal life histories in addition to focusing on the healthy and pathological varieties and gradations of consciousness as a complex phenomenon in nature. Naturally, my specific proposals and theoretical sketches may turn out to be wrong, but it is only in attempting to integrate evolutionary and ecological thinking with the science of consciousness that we can truly move towards a study of consciousness as a widespread natural, rather than merely human phenomenon. And to provide such a possibility proof for a bottom-up approach and evolutionary framework that can help us to think about the *raison d'être* consciousness has for organisms within their natural lives is the goal of this thesis. Finally, let me offer an outline for this thesis.

Outline of the Thesis

This thesis consists of six substantive chapters that each successively build on what came before.

Chapter 1 'A Darwinian Philosophy for the Science of Consciousness' provides a lot of the groundwork for the project of this thesis. Here, I will defend in more detail the idea that any biological approach to consciousness must address the question of what consciousness in all of its gradations and varieties does for healthy agents within the natural environments in which they have evolved. Furthermore, in order to provide a Darwinian standard for the field and advance a true biological science of consciousness, I will introduce and motivate my own hypothesis regarding the evolutionary origins and function of consciousness in hedonic evaluation that I dub the *pathological complexity thesis*, before explicating it in the following chapters to make sense of the place of mind in nature.

Chapter 2 'The Explanandum: Animal Consciousness and Phenomenological Complexity' has an important dual role. Firstly, I will show that any theory of consciousness must account for its varieties and gradations both within and across species, i.e. the explanandum of this thesis: *phenomenological complexity*. Secondly, I will tackle one possible objection to my project - the claim that since human consciousness is supposedly the only one we can access, that it will be impossible to investigate the phenomenological complexity of animal consciousness. Drawing on a recent call by Birch, Schnell, and Clayton (2020) to develop multi-dimensional frameworks for animal consciousness, I will show that we already have a wide range of experimental paradigms that allow us to differentiate and measure five dimensions of phenomenological complexity, and this will allow us to move towards a comparative study of consciousness that can ultimately remove humans from the centre of reference.

By breaking the features of consciousness down into the five different dimensions of self-consciousness, evaluative experience, sensory experience, the integration of

experience at a time, and the integration of experience across time, we will be able to substantially narrow the explanatory gap. It is here that Chapter 3 'The Origins of Consciousness or the War of the Five Dimensions' will further advance my goal of a bottom-up model of consciousness by trying to reverse-engineer the origin and function of consciousness - the dawn of 'qualia' - by making these dimensions face off against each other to determine the most probable candidate for the first sparks of subjective experience: evaluative experience.

Having motivated my epistemic bet on hedonic evaluation as the most basic kind of subjective experience, Chapter 4 'The Explanans: Pathological Complexity and the Dawn of Subjectivity' will attempt to develop the beginnings of a theory by explicating the pathological complexity thesis through the resources of physiology, neuroscience, ecology, evolutionary biology, and even economics. By making sense of the lifestyle changes of animals preceding the Cambrian explosion, we will be able to explain the dawn of subjective experience in the form of a basic feel of evaluation. The evolution of animal agency required an efficient solution to the problem of handling a body with high degrees of freedom that led to an explosion in pathological complexity. This design challenge led to the evolution of hedonic valence for efficient decision-making, and thus the origins of sentient beings - or as I shall call them, 'Benthamite creatures', in reference to the founder of utilitarianism Jeremy Bentham.

Having constructed an alternative theory of the origins of consciousness based in a model of evaluation, Chapter 5 'Pathological Complexity meets Phenomenological Complexity' will in turn develop phenomenological complexity as a response to pathological complexity, by showing how my theory can account for the four dimensions we have previously shelved off. The pathological complexity thesis will enable us to make predictions regarding the likely subjective mental states of animals based on the evolutionary history and ecological life-history challenges faced by them *in nature*, which can then be tested through various experimental means discussed in Chapter 2. The two groups I will give the most attention here are gastropods, as a rich source of evidence for the pathological complexity thesis, and arthropods, as a potential challenge. But I will also offer a discussion of octopuses, fish, non-avian reptiles, and corvids to show how the pathological complexity thesis can help us to scientifically approach the question of what it is like to be them.

Finally, Chapter 6 'Steps Towards the Final, Crowning Chapter of the Darwinian Revolution' will conclude this thesis by returning to Griffin's goal of moving us towards a Darwinian study of consciousness and examine how far the pathological complexity thesis has brought us towards this goal.

Chapter 1

A Darwinian Philosophy for the Science of Consciousness

> *Most of Darwin's basic ideas about evolution are now generally accepted by scientists, but the notion that there has been evolutionary continuity with respect to conscious experiences is still strongly resisted. Overcoming this resistance may be the final, crowning chapter of the Darwinian revolution.*
>
> – Donald Redfield Griffin (1998, p. 14)

1.1 Introduction

This thesis is a philosophical contribution to the emerging science of animal consciousness. It is a science that the prominent American ethologist Donald Griffin tried to establish in the 1970s when he called for a 'cognitive ethology', but which only truly began to shape into a genuine interdisciplinary field a decade after his death, with the 'Cambridge Declaration on Consciousness' in 2012 and the formation of the first interdisciplinary journal of non-human consciousness in 2015, aptly titled 'Animal Sentience'. As per the epigraph at the beginning of this chapter, the goal of this thesis is to advance Griffin's vision of the *final, crowning chapter of the Darwinian revolution* by helping this burgeoning field to cast off the chains of a pre-Darwinian view of the mind in both philosophy and science. This will allow us to transition towards a true Darwinian science of consciousness in which the evolutionary origin, function, and phylogenetic diversity of consciousness are moved from the field's periphery to its very centre and enable us to endogenize consciousness into an evolutionary view of life.

In the introduction to this thesis, I have emphasized that there are two senses in which we can ask about the place of consciousness in nature: one concerns the presence and contents of minds in nature, the other the relationship between matter and mind. The former has been the called the *problem of other minds*,[1] the latter is the familiar *mind-body problem*; two problems to which answers vary so incredibly widely that many are under the impression that there is no standard according to which we could even begin to sort out which views are likely going to be wrong. It is in this context that one may be forgiven for thinking that the historical question

[1] See Avramides (2020) for an elegant *Stanford Encyclopedia of Philosophy* article on the topic.

of how consciousness evolved constitutes anything but an additional problem that only further complicates the picture. Yet, it is precisely in the modern twenty-first century theory of evolutionary biology that we find the much-needed standard that the science of consciousness was so desperately lacking.

In order to develop a true biological science of consciousness, we must attend to the (cognitive) ethologist's demand to address the *functionalist question* of what consciousness, in all of its diversity and gradations, does for healthy agents within their normal ecological lifestyles and the natural environments they have evolved in. Accordingly, this thesis has two objectives: (i) to argue for the need for and possibility of an evolutionary bottom-up approach that addresses the problem of consciousness in terms of the evolutionary origins of a new ecological lifestyle that made consciousness worth having, and (ii) to articulate a thesis and beginnings of a theory of the place of consciousness as a complex evolved phenomenon in nature. This thesis can be succinctly summarized as follows:

Pathological Complexity Thesis:
The function of consciousness is to enable the agent to respond to pathological complexity.

Inspired by Godfrey-Smith's (1996a) *environmental complexity thesis* that sought to establish a link between environmental complexity and the evolution of cognitive complexity; "The function of cognition is to enable the agent to deal with environmental complexity" (p. 3), the pathological complexity thesis is grounded in the idea that health and consciousness are two closely related natural phenomena.[2] I will argue that the origin and function of consciousness lies in the capacity to help complex but vulnerable animals to deal with their species-specific health challenge to seek out the beneficial and avoid the pathological. Furthermore, I shall argue that naturalist understanding of this 'biological normativity' requires the development of a Darwinian theory of the organism that will in turn allow us to make sense of organisms as active *agents* and *subjects*. This will include their subjective experience as an integral part of our biological understanding of what makes a bat a bat, a snake a snake, and a healthy bee a healthy bee.[3]

Pathological and Phenomenological Complexity

The pathological complexity thesis is intended as a functionalist alternative to the false dilemma between the two dominant traditions in the philosophy of mind and the science of *human* consciousness, i.e. between strongly externalist representationalist theories of consciousness that overemphasize sensory experience, and strongly internalist ones that overemphasize self-awareness as the models for all of experience.[4] Instead, the pathological complexity thesis seeks to develop an alternative model of consciousness based on a model of *animal sentience*.

Because of the associations of the term 'consciousness' with the complexity

[2]In Chapter 4, I will discuss Godfrey-Smith's thesis and its connection to mine in more detail.

[3]I borrow this phrasing from an important quote of one of the 'fathers of ethology', Konrad Lorenz, that will be discussed shortly in this chapter.

[4]As will become clear throughout this thesis, this move is inspired by similar dilemmas faced by Darwin and the early ethologists, in addition to recent work from Peter Godfrey-Smith (see Marshall and Godfrey-Smith 2014).

of the human mind, the term 'sentience' - coming from the Latin verb *sentire*, i.e. 'to feel' - is often preferred among those with a primary interest in animal consciousness.[5] The term has not received universal endorsement, however, because it is often used ambiguously as (i) a deliberately broad and inclusive concept to refer to all kinds of subjective experiences, (ii) a reference to the most minimal kind of subjective experience found at the evolutionary origins of consciousness, or (iii) the hedonic capacity to feel pleasure or pain. Here, we can avoid these ambiguities because this thesis will combine all three interpretations. The origins and raison d'être of minimal consciousness or 'qualia' lie in hedonic evaluation as 'valence' (rating experiences as good, neutral, or bad). Sentience in this evaluative sense is an inherently 'interactionist' - or perhaps better, 'dynamic' - dimension of consciousness.

Pathological complexity is neither an internalist nor externalist measure, but emerges dynamically from the interaction of organism and environment as a measure of the complexity of an organism's life-history strategy, and will hence vary with the different 'lifestyles' of different animals. It can be understood as the computational complexity of the Darwinian trade-off problem faced by all biological agents as they deal with challenges and opportunities throughout their life-histories in order to maximize their fitness. As I shall argue in this thesis, consciousness evolved in the Cambrian explosion alongside a new evaluative animal *lifestyle* characteristic of large parts of the Metazoan branch of life, as an adaptive response to a computational explosion in just this kind of pathological complexity that made sentience worth having.

Importantly, I want to emphasize that I am not arguing that the complex trade-offs organisms are faced with are themselves pathological. I use the term 'pathological complexity' instead of the equally adequate and perhaps less confusing terms 'teleonomic complexity' and 'life-history complexity', not because I want to make the argument that organisms with greater life-history complexity are less healthy, but because I want to emphasize that it is only in understanding life-history tradeoffs that we can distinguish healthy from pathological trait variations of traits and that includes variations of consciousness both within and across species, i.e. what I call 'phenomenological complexity'. From an evolutionary perspective, we can understand health health as an optimal design-response to the species-specific pathological complexity trade-offs faced by different organisms in their life-histories. This means that health and pathological states come in degrees, rather than is typical in our human folk thinking about these states as something like a binary distinction. Indeed, I want to emphasize that there is nothing that must inherently follow from an evolutionary understanding of health for our ordinary practice of human medicine. I am here merely interested in a biological kind of normativity that is shared by all organisms and may be very different from how we want to think about 'health' and pathology when it comes to normative questions about which states ought to be treated in our own species. Indeed, I believe that it is precisely due to the ambiguity in how people usually think about the notions of health and pathology, that some have responded to my thesis that it could be expressed without making reference to these notions. Yet, in a naturalist investigation of health and consciousness, I shall argue that we should not make too much out of the human case. It is a special case

[5]See Browning and Birch (2022) for a recent review of animal sentience research.

of a much more widespread phenomenon we find in nature.[6]

Furthermore, unlike other theories of consciousness that struggle to make testable predictions, the pathological complexity thesis offers us a conjectural empirical framework for the relationship between mind and life by linking properties of phenomenological complexity - such as sensory experience, self-awareness, hedonic feelings, points of view, and mental-time travel - to properties of pathological complexity. A deeper understanding of what makes varieties of consciousness healthy and pathological will be of utmost importance for extending the Darwinian revolution towards consciousness, which is why parts of this chapter will be dedicated to explicating health as a natural phenomenon.

Finally, operationalizing pathological complexity in terms of the number of parameters and constraints in the evolutionary optimization problem studied by state-dependent or state-based behavioural and life-history theory offers us an elegant framework to naturalize health, organisms, and the idea of different ecological lifestyles central to a Darwinian approach of life and mind. It is my hope that the thesis of this thesis will provide us with a fruitful hypothesis and framework to move us closer towards a comparative science of animal consciousness that can help us to make sense of the place of mind in nature.

Chapter Outline

This chapter is organized as follows. Section 1.2 'Some Preliminary Remarks on Organisms, Health, and Philosophical Method' will offer some meta-philosophical reflections about naturalist philosophy and how we can make progress, despite resistance, to understanding notions such as health and consciousness from a Darwinian point of view. Section 1.3 'Lessons from the Darwinian Revolution', will defend an extension of the Darwinian program towards animal consciousness, by placing it in its historical, methodological, and social context. The history of the Darwinian revolution for biology and psychology will offer a number of important scientific and philosophical lessons and building blocks for the pathological complexity thesis that will accompany us throughout this thesis. Section 1.4 'Carrying Darwinism to Completion' will combine the foregoing lessons from the Darwinian revolution with modern state-based behavioural life-history theory to build the beginnings of a theory of health, organisms, and ecological lifestyles grounded in *pathological complexity* that can be used to endogenize consciousness into the Darwinian revolution.

1.2 Some Preliminary Remarks on Organisms, Health, and Philosophical Method

While my suggestion to link health and consciousness in terms of an association between the biologically normative properties of pathological complexity and the phenomenological complexity of organisms will be intuitive to many animal consciousness researchers focusing on sentience and evaluation, many contemporary philosophers may find it strange. Just as consciousness constitutes perhaps the core problem in the philosophy of mind, the proper definitions of health, pathology,

[6]Within the scope of this thesis we may even treat health in humans as a distinct notion entirely.

and normal functioning constitute the fundamental problems in the philosophy of medicine. This is plausibly part of the reason why philosophers of mind have been so reluctant to seriously consider evaluation - an inherently normative notion - as the most basic form of consciousness. After all, how could one hope to address one of the biggest problems in philosophy by solving another seemingly unrelated problem in an entirely different field?

An immediate concern should be that it is one thing to aim for progress on one of the major philosophical debates, but it is quite another to undertake the task of making progress on two of its most disputed controversies. Furthermore, the idea that these vexing phenomena share a intimate, but yet, unexplored connection may strike some more traditionally inclined philosophers as a bizarre project. Philosophy in the eyes of those within the tradition of Bertrand Russell's decompositional style of analytic philosophy is engaged with the detailed and narrow, rather than with the general and broad, but this is not the vision of philosophy that I endorse here.

A Darwinian Philosophy of Nature

This thesis follows a particular naturalist style of doing philosophy that is common in the sphere of Australian philosophy of biology and psychology. It is advocated by Antipodean philosophers such as Kim Sterelny, Paul Griffiths, and Peter Godfrey-Smith, but also embodied by the likes of Dan Dennett and Ruth Millikan, in which the biological sciences - and in particular modern evolutionary theory - become an instrument for the materialist philosopher: "a lens—through which we look at the natural world" (Godfrey-Smith 2013b, p. 4). Godfrey-Smith (2013b) has called this activity 'philosophy of nature' to reflect the older ambitions of the German tradition of *Naturphilosophie* to combine science and philosophy to make sense of the world and our place in it, though without excess metaphysical speculation, which nicely describes the project of this thesis.

The intended task of the philosopher here is to synthesize, rather than to analyze, the products of the sciences in order to construct better theories and models; which brings it in many ways indistinguishably close to the kind of integrative work done by important scientific names, such as Darwin himself. Indeed, this thesis is very much grounded in the idea that we can endogenize consciousness within modern evolutionary biology by following in the footsteps of Darwin and his followers. But my motivation is not merely scientific, it also has a distinctive philosophical flavour, that was beautifully expressed by Richard Rorty who described philosophy as being in the unique position of providing the only "place in the university where a student can bring any two books from the library and ask what, if anything, they have to do with each other."[7] This comparative and integrative ambition is very much the spirit of this big-picture thesis on the connection between pathological and phenomenological complexity. But before we can even begin to instigate such an investigation we first need some conceptual grip on the nature of our target phenomena.

What is Health and Consciousness?

One immediate philosophical problem for any biological investigation of consciousness and health is that the terms 'consciousness' and 'health' are notoriously ill-

[7]Quote attributed to Rorty in Godfrey-Smith (2013a, p. 4).

defined. The cognitive ethologist Frans De Waal (2016), for instance, notes that he prefers "not to make any firm statements about something as poorly defined as consciousness. No one seems to know what it is" (p. 23). Former zookeeper and animal welfare expert turned philosopher, Heather Browning (2020b), similarly expressed skepticism that health reflects "any naturally existing state", instead of a mere cluster of different phenomena (p. 164). If they are right, the pathological complexity thesis seems to rest on shaky ground; built to connect two phenomena that may not even exist.

But the absence of precise definitions for either should not stop us in our tracks. Both terms - as used by the public - may be vague, ambiguous, and resistant to the analytical philosopher's ideal of conceptual analysis. Indeed, if one's goal is to provide a definition of the term that would cover its varied usages, one may be tempted to conclude that we would be better off eliminating their folk concepts altogether.[8] But my goal is not conceptual analysis, it is *conceptual explication* (Carnap 1950) or as I have called it elsewhere, *naturalist conceptual engineering* (Veit and Browning 2020b), i.e. to construct concepts to better understand natural phenomena in light of empirical data from the sciences rather than merely appeal to intuition. We are trying to capture a phenomenon in nature, for which the neurophilosopher Patricia Churchland (2002) suggests that we should simply rely on common sense to establish "provisional agreement" on a number of "unproblematic examples of consciousness" (p. 133). There is no need to provide a philosophically satisfactory concept of consciousness or health before we can begin to investigate them, any more than we would need to define the concept of koala before we can learn about their enjoyment of eucalyptus leaves. A precise definition that could be used to demarcate controversial cases is the endpoint of a naturalist investigation, not its starting point. More importantly, our folk understanding of consciousness - which is almost exclusively based on intuitions about our own human case - may be radically revised in light of scientific data and investigation of other animals. My hope here is to develop a better understanding of consciousness, its origin, and its place in nature as widely shared trait in the animal kingdom by drawing on the sciences and in particular evolutionary biology. This issue will come up especially in Chapter 3, when I try to reverse-engineer the origins of consciousness by breaking the phenomenon into different components and offer a partial revision of the typical philosopher's concept of consciousness by drawing on experimental philosophy research into how ordinary people conceptualize consciousness.

Scientists repeatedly proceed to investigate phenomena that have so far remained elusive, proving that vagueness need not be an obstacle to scientific inquiry (Neto 2020). In this naturalist activity, it is ultimately nature, not intuition, that will decide how we should understand consciousness, precisely because as Figdor (2018) argues, we "lack widely accepted theories and models that can organize and articulate the pre-theoretic consciousness-related concepts we are using to guide our initial investigations" (p. 10). Following Churchland (2002), we can confidently reply that we can at least initially "use the same strategy here as we use in the early stages of any science: delineate the paradigmatic cases, and then bootstrap our way up from there" (p. 133).

Paradigmatic cases of consciousness there are plenty: pain, pleasure, smell,

[8] Wilkes (1984), for instance, has argued so in the case of consciousness and Hesslow (1993) and Ereshefsky (2009) in the case of health.

vision, taste, a sense of one's body, memories, alongside a whole other range of subjective experiences. Similarly, we have some intuitive grasp of health and pathology in humans and animals alike, such as diseases, broken bones, lesions, parasites, burns, poisons, maladaptive behaviour, and other 'biological wrongs' - even if we have struggled to derive something like a folk theory of health. So it is perhaps unsurprising that we can also intuitively distinguish healthy subjective experiences from unhealthy ones such as major depressive disorder, anxiety disorder, aphantasia, synesthesia, autism, schizophrenia, prosopagnosia, chronic pain, and many more. Yet, many philosophers of medicine would *deny* that health is natural phenomenon, thus perhaps providing an explanation for why philosophers have given so little attention has been given to the search for the origins of consciousness in the normative notion of evaluation.

Resistance to Naturalism in the Philosophy of Medicine

Despite naturalist views being discussed by philosophers of medicine, their assessment is largely negative, and most within the field now maintain that health reflects personal evaluations or the values of society at large; a consensus that health is primarily a *normative* concept, rather than 'only' an objective *biological* property of organisms.[9]

Such a view may have been the dominant one ever since the French historian of science and first modern philosopher of medicine, Georges Canguilhem (1991) argued in his influential treatise *The Normal and the Pathological* that "[t]here is no objective pathology. Structures or behaviors can be objectively described but they cannot be called ['']pathological['] on the strength of some purely objective criterion" (p. 226). Others like Lennart Nordenfelt (1995), who emphasize the concept of agency, have argued that health cannot be understood in a reductionist naturalist way and instead requires a more holistic conception, where it is understood as the ability to achieve one's vital goals. Phenomenologists such as Havi Carel (2007) have similarly argued that the "experience of illness cannot be captured within a naturalistic view" (p. 95). Such strong assertions against the very *possibility* of a naturalist account are surely premature and yet can be found throughout the literature, effectively making naturalism a 'boogeyman' of the field. Rarely has there been a philosophical debate in which naturalism has been so forcefully and unceremoniously dismissed.

This anti-naturalist consensus in the field can be usefully summarized as an appeal to the 'irreducibility' of (i) the normativity of health and disease, (ii) the loss of agency in health and disease, and (iii) the phenomenology or subjective experience of health and disease. But who is to deny that these features can be part of a naturalist account of health and disease? Naturalist philosophers have long worked on attempts to make these notions of normativity, agency, and phenomenal experience safe for naturalism.[10] What all of these anti-naturalists curiously share, though not necessarily all other philosophers of medicine opposed to a naturalist view of health, is an emphasis on *subjectivity*. Similarly to those who view naturalist explanations of consciousness as deeply problematic, they argue that the very idea

[9] See Dominic Murphy's (2020) SEP article "Concepts of Disease and Health" for an excellent recent overview.

[10] Note the mirroring of the title of this thesis.

of a naturalist account of health and disease is mistaken. They hold that one cannot account for health and disease from the objective third-person perspective of science, since they are phenomena at the level of a *subject*, not an *object*, and science cannot account for the former - a view familiar from so-called 'naysayers' who assert a scientific account of consciousness to be impossible.[11]

This way of thinking about naturalism, however, is highly problematic. Subjects aren't some mysterious entities inaccessible to science: they are an evolutionary product and also include non-human animals. But the possibility of a Darwinian reconciliation between a view of health as a property of the organism as an 'object' and of the organism as a 'subject' has been given scant attention, precisely because *non-human health* has been less than an afterthought in this debate (see Matthewson and Griffiths 2017).

As I will argue in this chapter, health and pathology are not only perfectly naturalistic concepts, but they play important roles in evolutionary biology, and will help us to extend the Darwinian revolution to include consciousness.

1.3 Lessons from the Darwinian Revolution

> *The difference in mind between man and the higher animals, great as it is, certainly is one of degree and not of kind.*
>
> – Charles Darwin (1871, p. 105)

The first chapter of this thesis is titled 'A Darwinian Philosophy for the Science of Consciousness' precisely because the so-called emergence in the 1990s of a science of consciousness has at best been a science of *human* consciousness. From a naturalist perspective, we can only truly claim to have established a Darwinian science of consciousness once we study consciousness as a natural, rather than a human, phenomenon - and this must include all sentient beings. Unfortunately, consciousness appears to be one of the last biological phenomena that we have failed to integrate into the Darwinian revolution. By this revolution, I am referring to Darwin's theory of natural selection that fundamentally changed our view of life and our place among other organisms. Just like humans are no longer seen as the top on a hierachical ladder of creation, but instead only one branch among many other species in an egalitarian tree of life, I argue that our study of consciousness must dethrone humans from the centre of attention and study them as a mere special case. Consciousness is older and much more widely spread than an anthropocentric picture suggests, but to realize this we will have to seriously investigate its plausible evolutionary origins.

As I noted in the introduction to this thesis, such attempts to extend the Darwinian revolution have been burdened with unfortunate epithets such as 'just-so stories', demonstrating a resistance to the possibility of an adaptationist evolutionary process explanation of how the mind gradually came into existence. Lewontin (1989) himself, who has been one of the fiercest opponents of adaptationist explanations in biology,[12] contributed to this skepticism when he argued that we know next to nothing about the evolution of the human mind and probably never will. Such a pessimistic attitude is certainly not entirely unfounded, since consciousness appears

[11] Flanagan (1991) offers an excellent critique of these consciousness skeptics.

[12] See Lewontin (1979, 1981); Gould and Lewontin (1979); Levins and Lewontin (1985).

to leave no fossil trace, making it seemingly impossible to trace its phylogenetic origins and reconstruct its raison d'être through a historical narrative explanation.

In this, however, consciousness is not alone - sharing a fate with a wide range of other complex biological phenomena that people thought could not be explained in Darwinian terms. Most notably of these is behaviour, which has been firmly integrated into modern evolutionary biology since the ethologists endogenized it within Darwin's explanatory framework. Paying close attention to the origins of Darwinism and its extension to behaviour will provide a number of useful lessons for a *cognitive ethology*, that likewise endogenizes consciousness in a Darwinian view of life.

1.3.1 Darwinism and Teleonomy

In trying to provide a Darwinian account of consciousness we have to clarify what we mean by such a project. Above, I noted that the pathological complexity thesis rests on the Darwinian idea of a functionalist alternative to a false dilemma between externalist and internalist approaches to consciousness. Internalist explanations seek to explain features of a system in virtue of other features of that system - of processes, structures, organization, and development *within* it, rather than outside of it. Externalist explanations, on the other hand, aim to explain features of the system by recourse to the external, i.e. the environment - Godfrey-Smith (2002) calls them 'outside-in' explanations (p. 30). This distinction is not only relevant for categorizing different views of the mind, but also of life itself, since many treat Darwinism (mistakenly) as an externalist program.

Lewontin has stated the alleged link between Darwinism and externalism perhaps the most forcefully, arguing that the success of the Darwinian project was due to its disentangling of internal and external forces that have previously been inseparable.[13] Darwin broke with what Lewontin called *transformational* theories of the past, such as Lamarck's (1984) theory of evolution that postulated change to individuals within their life-histories arising from 'subjective' or what we may want to call 'internal' forces, such as will and striving. The Darwinian theory of the organism made it the "*object*, not the subject, of evolutionary forces" such as natural selection and random drift that are "autonomous and alienated from the organism as a whole" (1985, p. 85). To complete the Darwinian revolution, however, Lewontin maintained that the internal forces - the subject-side of organisms - must be re-introduced:

> Darwinism cannot be carried to completion unless the organism is reintegrated with the inner and outer forces, of which it is both the subject and the object.
>
> – Richard C. Lewontin in (Levins and Lewontin 1985, p. 106)

By this, Lewontin did not mean subjective experience, but rather how organisms as agents actively 'participate' in their evolutionary path and 'construct' their environments, as an alternative to a traditional adaptationist view of life. These notions of agency and construction have been highly influential in modern attacks on Darwinism (Ho and Saunders 1979; Laland et al. 2014; Noble 2015; Müller 2017), but I am not here interested in the conceptual role of organisms as subjects for challenging

[13] See also Lewontin and Levins (1997).

the theoretical modeling of evolution. My interest lies in subjects as an evolutionary product, to allow us to make sense of the evolution of subjective experience. As Godfrey-Smith (2017c) notes in his discussion of Lewontin, subjects are not only a cause of evolutionary change, they are also its product.

In advancing a gradualist view of the evolution of consciousness, theoretically less loaded terms like 'agency' and 'subjectivity' are useful for thinking about organisms as being more or less subject-like; they "can realize subjectivity to a greater or lesser degree" (Godfrey-Smith 2017c, p. 1). While subjectivity may appear similarly as elusive as consciousness, it does not similarly suffer from an overabundance of theoretical frameworks. We can, as Godfrey-Smith (2019a) argues, use Lewontin's distinction between objects and subjects to bridge the gap between matter and mind: "[t]he history of life includes the history of subjectivity, and subjective experience is the experience *of a subject*" (p. 2). And in doing so we may be able to carry the Darwinian revolution to its completion.

Unlike Lewontin, however, I do not see a conflict between adaptationism and an explication of the subject-side of organisms. As this thesis hopes to demonstrate, it is precisely with a Darwinian view of organisms that we will be able to make sense of 'subjectivity'. This does not mean that we can't recognize that evolutionary biology has been dominated by externalist modes of explanation, with features of the organism being explained in terms of their adaptive fit to their external environment (Godfrey-Smith 1996a; Walsh 2015). Evolutionary biologists readily admit that "[t]he suspicion of internal causes in the dominant neo-Darwinian culture ran so deep that every internalist idea, no matter how reasonable, was treated as an appeal to vitalism" (Stoltzfus 2019, p. 46). But we should distinguish the idealization choices made by *some* modellers, from a deeper commitment to the necessity of an externalist view of adaptations. After all, there is plenty of modelling work done by evolutionary biologists that can be seen as 'internalist', such as the study of game theoretic dynamics that emerge from the structure of a population rather than its external environment (Sterelny 1997, p. 556). Indeed, it is a mistake to think of adaptationism and externalism as a one-package-deal. As I shall argue, we can straightforwardly follow Sterelny's (1997) suggestion to decouple adaptationism from externalism and consider the two separately.

Many of the arguments against adaptationism are really arguments against its externalist versions that use a so-called 'lock and key' model of the adaptation between organism and environment; a criticism that need not apply to other versions. Modern evolutionary biology recognizes plenty of feedback between organisms and the species-specific environments in which natural selection takes place, such as Brandon's (1990) notions of 'selective environments' and 'ecological environments', which can be distinguished from an organism-neutral externalist view of the environment. The external features that *matter* to the evolutionary trajectory of the organism are themselves causally dependent on the organism. No longer do modern evolutionary biologists see adaptations in the externalist design-sense of a natural theologian such as Paley (1802), who argued that animals are a proof of God's design plan, with species being fitted to preexisting external niches.

As with many scientific concepts, the concept of adaptation came to be redefined - or rather explicated - in a naturalistically unproblematic sense referring to whatever is produced by natural selection, even if such 'design' appears inefficient and wasteful (Griffiths and Gray 2001, p. 209). Much of the opposition from 'Neo-Darwinians' to

Gould's and Lewontin's criticism of 'adaptation' was based on a mismatch between a usage of that term in its original pre-Darwinian sense and its modern explication, which already included at least some of the features of feedback between organism and environment that were alleged to be lacking in the modern Neo-Darwinian view of life. Instead of seeing Darwinism as an externalist theory of organismal traits that replaced previous vitalist modes of thinking that were confused between internal and external forces, we should see it as a *teleonomic* rejection of a false dilemma between internalist theories such as Lamarck's and a strongly externalist view of organisms being designed by a benevolent God to fit their environments, by providing us with an inherently 'dynamic' or 'interactionist' picture of the living world.

By 'teleonomic' I am employing Pittendrigh's (1958) coinage of the term, as a naturalistically unproblematic Darwinian replacement for older and mistaken teleological notions about the purposefulness, design, and normativity of life, which is why I noted above that my notion of 'pathological complexity' could alternatively have been called 'teleonomic complexity'. By understanding organisms as goal-directed systems or Darwinian *agents* evolved to maximize their fitness, our understanding of health, just like our understanding of adaptation and design will come to be transformed. As I shall argue in this thesis, we can build a theory of the organism as both an object and *subject* with the tools of modern state-based and behavioural life-history theory, which does not - as Lewontin objected to - treat organisms as machines with mosaic-like traits, but rather as agents having to deal with integrated bundles of trade-offs in organismal design. It is precisely this teleonomic theory that will bring out the subject-side of organisms. With this, let us now turn to Darwin's own speculations about the evolution of mind.

1.3.2 Early Darwinian Views of Mind

As the approach in this thesis is inspired by Darwin, it will be hardly surprising that Darwin himself rejected a dualist view of the mind, both in a metaphysical sense and in the phylogenetic sense of a sharp dividing line between us and other animals. Following the success of his 1859 book *On the Origin of Species* he published two further very influential books in which he sought to defend a continuity view between us and other non-human animals. In his *The Descent of Man, and Selection in Relation to Sex*, Darwin (1871) argued that "the lower animals, like man, manifestly feel pleasure and pain, happiness and misery" (p. 39) and that "there is no fundamental difference between man and the higher mammals in their mental faculties" (p. 35). And in his *The Expression of the Emotions in Man and Animals*, Darwin (1872) went on to vastly expand his hypotheses on the evolution of mind, in particular the emotions.

But while Darwin urged us to think about the mind in terms of evolutionary continuity, he deliberately avoided public speculation on the very origin of mind and life, noting: "I must premise that I have nothing to do with the origin of the primary mental powers, any more than I have with that of life itself" (1859, p. 207). Yet, it is clear that the evolution of consciousness sincerely troubled him and some of his early notes revealingly contained the questions: "How does consciousness commence?" and "Where pain & pleasure is felt where must be consciousness???", suggesting that even Darwin speculated about the origins of sentience.[14] Later, he

[14]See Darwin's *Old & useless notes about the moral sense & some metaphysical points* (1840,

repeated his resistance to explaining the origins of life and mind as problems that ought to concern us *now*: "In what manner the mental powers were first developed in the lowest organisms, is as hopeless an enquiry as how life itself first originated. These are problems for the distant future, if they are ever to be solved by man" (1871, p. 36). But as with the origins of life, early Darwinists were immediately spurred on to think about the origin of mind.

Indeed, the idea of thinking about the mind as a product of evolutionary forces immediately influenced important figures such as Herbert Spencer, Thomas Henry Huxley, Ernst Heinrich Philipp August Haeckel, George John Romanes, William James, Conway Lloyd Morgan, James Mark Baldwin, and John Dewey, who all substantially contributed to an early evolutionary understanding of the mind.[15] What we saw in the decades after Darwin, Ginsburg and Jablonka (2019) note, was that "all psychologists, philosophers, and biologists who considered mental evolution and the evolutionary origins of mentality explained it in terms of natural selection" (p. 71). Indeed, it was common at this time to think that the mysteries of the mind could be unveiled by viewing them through an evolutionary lens.

Spencer, for instance, insisted that "[i]f the doctrine of Evolution is true, the inevitable implication is that Mind can be understood only by observing how Mind is evolved" (1870, p. 291). Moreover, both Dewey and Spencer endorsed a continuity thesis between life and mind that influenced this thesis here: the mind is seen as the natural consequence of the evolution of complexity.[16] Furthermore, Romanes speculated in some detail that pleasure and pain may be the key to understanding the place of consciousness in nature:

> Possibly, however—and as a mere matter of speculation, the possibility is worth stating—in whatever way the inconceivable connection between Body and Mind came to be established, the primary cause of its establishment, or of the *dawn of subjectivity*, may have been this very need of inducing organisms to avoid the deleterious, and to seek the beneficial; the raison d'être of Consciousness may have been that of supplying the condition to the *feeling of Pleasure and Pain*.
>
> – George John Romanes (1883, p. 111) [italics added for emphasis]

Evolutionary thinking naturally lends itself towards a view in which sentience constitutes the origin of consciousness. That organisms would evolve to value states and behaviours that increase their own fitness and avoid those that are detrimental to their health appears not at all mysterious from a Darwinian point of view.

Unfortunately, discussions of consciousness and its evolution, in both humans and non-human animals, went out of fashion in the early twentieth century. This was largely as a result of the rise of the behaviourist program coming from Watson and the more radical behaviourism of Skinner, who turned *American* psychology into the study of mere behaviour, banning consciousness from science. But while their official doctrine has been all but abolished, their influence in the study of consciousness remains alive and well. In the following, we will take a closer look at

p. 35).

[15] While Spencer (1855) had already written about the evolution of mind four years prior to Darwin's *On the Origin of Species* (1859), he embraced Darwin's theory of natural selection with open arms.

[16] See Godfrey-Smith (1996a) for a detailed discussion of their views.

the rise of behaviourism, classical ethology, and Griffin's eventual call for a cognitive ethology in order to understand what it means to take a truly Darwinian approach to life and mind.

1.3.3 Jamesian Psychology and the Rise of Behaviourism

To understand the rise of behaviourism one must understand the status of psychology at the beginning of the nineteenth century. One name that has perhaps influenced the science of consciousness more than any other is that of the aforementioned American philosopher and psychologist William James.

James is often credited for turning psychology into a discipline independent from philosophy with his 1890 textbook *The Principles of Psychology*, an achievement that made him the so-called 'father of American psychology' in the eyes of many. In the early development of psychology as a science, consciousness played an important role, so it should hardly be surprising that James is also praised as the "father of modern consciousness studies" (Ginsburg and Jablonka 2019, p. 41). Unfortunately, James had little to say about animal consciousness or its evolutionary origins, despite his interesting speculations about consciousness as the emergence of a new kind of evaluative agency and his emphasis on a functionalist view of the mind. His explanatory target was ultimately human consciousness, which he believed was undeniable, almost unique in kind, and could best be studied through the method of personal introspection.

The focus of psychology on consciousness, however, quickly came to be questioned. With the further development and success of psychological experiments, appeals to subjective states were less and less seen as "necessary to justify the value of experimental research" (Burghardt 1985, p. 914). The behaviourist program that tried to banish all mental concepts from psychology, and turn the science of the mind into a science of behaviour, can be seen as a natural outcome of this trend with functionalism coming to be abandoned. However, this did not mean that the behaviourists were anti-Darwinian, at least not initially. Indeed, unlike James who centered psychology around human consciousness, the behaviourists positively emphasized the importance of studying non-human animals due to their evolutionary continuity with us.

It is unfortunate that Watson, who is usually credited as being the father of the behaviourist movement, is often demonized and misdescribed. When attention is given to his early work, "presentations are usually brief and frequently contain a variety of errors" (Todd and Morris 1986, p. 71). Rather than treating behaviour as a black box, Watson showed a keen interest in the neurophysiology of animals, dissecting them with great care and experimental detail. However, after a decade of rigorous and methodologically diverse work on animal behaviour, Watson was ultimately fed up with having to justify the value of his research, after being repeatedly faced with the skeptical question of what his work could possibly teach us about human consciousness.

This should immediately remind us of the question not uncommon in twenty-first century *human* consciousness science and the philosophy of mind, regarding what the bearing could possibly be of work on animal consciousness. Watson's response to his detractors could hardly have been more Darwinian. In his 1913 paper "Psychology as the Behaviorist Views It" - the founding manifesto of the

behaviourist tradition that was meant to put these critics to rest - Watson explicitly defended the Darwinian view that there is "no dividing line between man and brute" (p. 158). To understand behaviour as a natural, rather than human phenomenon, he maintained was how we could only truly advance a science of behaviour as a natural phenomenon. Those inspired by Jamesian psychology, he harshly accused of being stuck in a pre-Darwinian mindset:

> [T]o make consciousness, *as the human being knows it*, the center of reference of all behavior, forces us into a situation similar to that which existed in biology in Darwin's time.
>
> – John B. Watson (1913, p. 124) [italics added for emphasis]

To understand the phenomenon of life, biologists readily recognized that an exclusive look at humans would lead to a biased picture, if not because of its complexity than because of the appeal to thinking of the human body plan as 'perfect' or 'higher' than other species. What was needed to truly revolutionize our understanding of life as a natural phenomenon, was an evolutionary approach based on phylogeny, the comparative method, and sound ecological thinking. Yet, early work in evolutionary biology was initially held back by its focus on the question of human descent.

This *starting point* was perhaps not surprising in a historical sense, since a sharp dividing line between humans and the rest of nature was considered to be the greatest challenge to Darwin's theory of natural selection. An assumption of human uniqueness had to be overcome. After the continuity between humans and apes was settled, biologists were finally able put humans in their place in nature, i.e. one among many species; man was dethroned. In trying to understand biological phenomena, biologists would henceforth use the comparative method - gathering evidence from many different species of animal and plants alike to learn general lessons about life. But from the perspective *of* Darwinism as a research program that placed us alongside rather than above all other life-forms, this early focus on humans must have seemed strange. As Watson (1913) put it: "Man ceased to be the center of reference" (p. 125).

In arguing against psychology as the 'science of the phenomenon of consciousness', Watson provided us with Darwinian arguments that very well apply against the top-down human-centric focus of the so-called 'science of consciousness' of today. But before we turn to Griffin's cognitive ethology as an attempt to develop a bottom-up biological study of the mind, let us first look at the classical ethologists in order to understand how they extended the Darwinian revolution towards behaviour.

1.3.4 Ethology, Health, and the Darwinization of Behaviour

In their introduction to the philosophy of biology, Sterelny and Griffiths (1999) define ethology as "the study of animal behavior under its normal ecological conditions (as opposed to unusual laboratory conditions) and from an evolutionary perspective" (p. 385). And this is certainly how many now think about it, as a tradition that was in opposition to the lack of ecological and evolutionary thinking shown by the behaviourists, and one that has now largely been superseded by behavioural ecology. But there was a more philosophical conviction that motivated its founders, one of of a teleonomic view of life, and this has largely gone unnoticed.

When one hears the term 'ethology', inevitably the names Konrad Lorenz, Nikolaas Tinbergen, and Karl von Frisch come to mind as the joint receivers of a Nobel Prize in Physiology or Medicine in 1973 for their involvement in the establishment of ethology and their discoveries of "organization and elicitation of individual and social behaviour patterns" (Nobel Prize Outreach 2021). Famously, Lorenz studied the imprinting behaviour of greylag geese, who show an innate instinct to bond with the first moving entity they encounter, whereas von Frisch was one of the first to study the 'waggle dance' used for communication by bees. Lastly, Tinbergen spent much of his time studying so-called 'fixed action patterns' such as the egg rolling of the greylag goose (see Beer 2020). But it is not their work on instincts and other proto-cognitive capacities that is of relevance to this thesis. Neither am I interested in their philosophical objections to the study of subjective experience. The reason we look here at the (classical) ethologists is the same reason we looked at Watson's motivation for the behaviourist manifesto: that is, to emphasize a Darwinian principle that motivated the origins of their approach.

Both the ethologists and the behaviourists wanted to establish an objective science of behaviour in which we rely on a bottom-up approach that emphasizes the study of simple behaviours in order to understand more complex ones. But the ethologists hardly saw the behaviourists as Darwinians at all. This is ironic, considering that both the (early) behaviourists and ethologists used Darwin to motivate their approach. However, we can readily resolve this puzzle. Whereas the behaviourists emphasized the alleged externalist explanatory style of Darwin's theory of natural selection, ethologists emphasized the theory itself, with its emphasis on function, survival value, and evolutionary phylogeny as sources of mechanisms to deal with the environments faced by organisms. This teleonomic perspective is nicely drawn out in a press release from the Karolinska Institute, which announced the Nobel Prize for the founders of ethology, and described their approach in a very Lorenzian manner as a Darwinian way out of a dilemma between the behaviourist's externalism and the vitalist's insistence on internalist forces:

> During the first decades of this century research concerning animal behaviour was on its way to be stuck in a blind alley. The vitalists believed in the instincts as mystical, wise and inexplicable forces inherent in the organism, governing the behaviour of the individual. On the other hand reflexologists interpreted behaviour in an one-side mechanical way, and behaviourists were preoccupied with learning as an explanation of all behavioural variations. The way out of this dilemma was indicated by investigators who focused on the *survival value* of various behaviour patterns in their studies of species differences. Behaviour patterns become explicable when interpreted as the result of *natural selection*, analogous with anatomical and physiological characteristics.
>
> – Nobel Prize Outreach (2021) [italics added for emphasis]

Following the end of the First World War, American psychologists came to treat appeals to instincts as unscientific explanations (Griffiths 2008). Instincts - in the sense of *unlearned* responses - were seen with unease, due to their teleological and internalist nature, their inability to yield themselves to physiological investigation and thus likewise causal explanation, in addition to their being tied up with mistaken purposive and vitalist conceptions of life (Dunlap 1919; Kuo 1921; Tolman 1923; Griffiths 2008). Whereas the likes of Darwin and James strongly endorsed the idea

of instincts, with the externalist turn of the behaviourists the notion came to be seen as unscientific. Lorenz, however, resisted this response as something that went too far in the opposite direction, maintaining that "it is hardly an exaggeration to say that the large and immeasurably fertile field which innate behaviour offers to analytic research was left unploughed because it lay, as no man's land, between the two fronts of the antagonistic opinions of vitalists and mechanists" (1950, p. 232). Indeed, Lorenz simply tried to do for behaviour what Darwin's naturalism had previously achieved for a similar false dilemma between vitalists and mechanists on the nature of life and organismic activity, i.e. to emphasize the teleonomic nature of animals:

> [B]oth mechanists and vitalists were incurably inhibited by quite specific conceptual errors and prejudices that were magnified by their clash of opinion. This prevented them from initiating research into animal and human behavior at the point where it should have begun, namely with straightforward, unprejudiced *observation* of healthy animals living under normal conditions. They were quite incapable of seeing behavior for what it is, that is, as an extremely complex, organic *systemic entity* consisting of quite different components; one which, like *any* organic system, owes its particular constitution to a quite specific historical process of development.
>
> – Konrad Lorenz (1997, p. 213) [italics in original]

To understand ethology as the mere opposition to exclusive laboratory work would be to miss out on this most important observation: ethology was intended as a teleonomic science (see also Thompson 1986b,a). In claiming to study learning in healthy organisms, the behaviourists did not recognize that it made no sense to think of health outside of the ecological context animals evolved in. As Lorenz (1981), who himself earned a Doctor of Medicine, put it: "the pathologic can be defined only by having recourse to ecological concepts" (Lorenz 1981, p. 57).

Unfortunately, this philosophical insight of the ethological tradition has come to be neglected, next to the more popular methodological slogan of studying animals in the wild, despite the fact that it was precisely their teleonomic reasoning that made them emphasize the importance of studying the lifestyles of healthy animals in their natural environments. This is why Lorenz praised Oskar Heinroth as the real founder of ethology as the comparative study of behaviour:

> Accordingly, a finely developed sensitivity to the delicate and *often diffuse* boundary between the not-quite-normal and the already pathological is perhaps the most important talent that a scientific animal keeper must possess! On the other hand, however, the keeping of animals in our sense is an apprenticeship that renders one's feel for the pathological just as acute as actual training in a medical clinic. It is surely no coincidence that Heinroth, the outstanding master of animal keeping as a scientific method, was, like the writer of this text, initially trained as a *medical doctor*.
>
> – Konrad Lorenz (1997, p. 229) [italics in original]

While the research of both Lorenz and Heinroth was admittedly often only conducted under quasi-natural circumstances, their work was informed by life history observations of healthy and pathological animal behaviour in its natural environment. This combination of observation and adaptationist thinking was inspired by

the descriptive natural history activity of Darwin and the taxonomic activities that made a distinction between healthy and pathological specimens in order to describe what a 'normal' species-typical individual ought to look like. This was a major component of how the Darwinian revolution began and this was thus how Lorenz thought a biological study of behaviour must get off the ground. It must be able to distinguish healthy from pathological behaviour, just as physiology and taxonomy have distinguished healthy from unhealthy phenotypes. This is why he emphasized early on that behaviour could be treated in just the same way as any other adaptive phenotype:

> What behaviorists exclude from the narrow circle of their interest is not only other learning processes, but simply everything that is not contained in the process of learning by reinforcement-and this neglected remainder is *neither more nor less than the whole of the remaining organism*! [...] What remains uninvestigated is all that makes an octopus an octopus, a pigeon a pigeon, a rat a rat, or a man a man, and, most important of all, what makes a healthy man a healthy man, and an unhealthy man a patient.
>
> – Konrad Lorenz (1981, p. 71) [italics added for emphasis]

Importantly, such a view is not in conflict with the observation by Ernst Mayr (1994) that the Darwinian revolution was at its core a shift towards "population thinking" from what he described as "typological thinking" prevalent in the biological sciences since Plato and Aristotle where organisms were treated as instances of an ideal type. Nevertheless, it is worth responding to the potential objection that if species are mere sets of diverse individuals hanging together through genealogical relationships, how could we possibly identify which organisms are 'healthy' or 'normal' and distinguish them from 'pathological' or 'abnormal' organisms?

While talk of 'normal' individuals within a population may be suggestive of a typological view of the organism, the concept has long been Darwinized. Indeed, we can easily distinguish the idea that there is an ideal 'healthy' or 'normal' type for each species from the idea that some *variation* within a population is 'normal' or 'healthy'. Rather than thinking about health here in terms of how close an organism is to an ideal type, we instead think about the variations within the population themselves and their role in an evolutionary view of the *population*. As Matthewson and Griffiths (2017) argue, we can define the normal (and distinguish the pathological) through reference to evolutionary norms, which can either be backwards-looking (are functions being performed that led to the evolution of the organism's traits) or forward-looking (does the organism perform functions that will allow its traits to persist in the population over time). Health - from an evolutionary perspective - can simply be seen as a fitness-optimum in the design space of the species, i.e. their evolved species-specific life-history strategy, and is thus entirely part of a Darwinian population view of life.

Finally, I would like to respond to the possible objection to my general approach drawing on the Darwinian revolution, that there hasn't really been such a scientific revolution in the nineteenth century. One historian who has perhaps the most forcefully argued against the idea of a great Darwinian revolution of biological science has been Bowler (1992, 2013), who maintains that Darwin's influence should be seen as more of a myth, since theories of evolution were abundant before him and science would have likely arrived at something like his views regardless. Yet, his argument

that Darwin's theory of natural selection only came to be accepted widely in the twentieth century with the *Modern Synthesis* between Darwin's views and those of Mendel is not at all a problem for the arguments I defend here. The Darwinian revolution still happened - even if it happened later than is commonly asserted. The extension of the Darwinian revolution towards behaviour in the twentieth century is perfectly compatible with such a latecomer view. Furthermore, Bowler does not sufficiently acknowledge the differences between the teleological views of evolution that were indeed prevalent at the time of Darwin, and Darwin's teleonomic naturalization of such thinking that can legitimately be seen as causing a teleonomic revolution - first in physiology, and later with the ethologists in behaviour - to naturalize teleological ideas of purpose, goal-directedness, function, and importantly health and pathology. Whether Darwin is solely responsible for these changes is beside the point. These notions gradually came to be conceptually re-engineered in the light of Darwin's theory of natural selection and it is because of this conceptual revolution that we can justifiably call speak of a Darwinian revolution (see also Richards 1992).

To conclude, the emphasis on the distinction between healthy and pathological states is an important lesson we can learn from the ethologist's Darwinization of behaviour and it is unfortunately a lesson the science of consciousness has not yet learned. If we are interested in making progress on the problems of consciousness within the next century, we must follow the ethologists' dictum to distinguish healthy from pathological variations of consciousness by thinking about their survival value in an ecological context. With these important lessons regarding the importance of health and the possibility of a teleonomic alternative to the dilemma between externalist and internalist explanatory schemes, let us now turn to Griffin's call for a *cognitive* ethology.

1.3.5 Donald Griffin's Call for a Cognitive Ethology

At the beginning of this chapter, I placed an obligatory epigraph by Donald Griffin who, beginning in the 1970s, tried to break the behaviourist's hold on the study of animal minds. I say obligatory because Griffin, in the last three decades of his life, has done more than any other scientist to re-establish and give credibility to a Darwinian study of consciousness. Admittedly, Griffin would not have had this impact if he hadn't discovered bat echolocation together with fellow Harvard undergraduate Robert Robert Galambos in 1938. When Thomas Nagel visited his institution from 1973 to 1974 and wrote up his famous paper 'What Is It Like to Be a Bat?', directly drawing on Griffin's discovery, Griffin felt immediately compelled to write up his thoughts on the possibility of a study of animal consciousness, resulting in his 1976 book with the fitting title *The Question of Animal Awareness: Evolutionary Continuity of Mental Experience*. Against Nagel, who denied the very possibility of learning about the subjective mental lives of other animals, Griffin (1976) argued that we could and ought to have a naturalistic investigation of the minds of other animals using the tried and tested ethological methods of naturalist observation in the wild, controlled experiments, and sound evolutionary reasoning - calling this thoroughly Darwinian discipline of animal minds 'cognitive ethology'.

But while Griffin managed to break the taboo against the scientific investigation of animal minds, cognitive ethology unfortunately never took off as a significant re-

search program on its own.[17] Worse, the term became for many a dirty word, associated with unscientific speculation, mere anecdotal evidence, and the *sin* of anthropomorphism (Allen and Trestman 2017). Here, it is important not to confuse 'cognitive ethology' with the general study of animal cognition (Allen 2004). After a long period of struggles, the subject of animal cognition finally received its own journal with the establishment in 1998 of *Animal Cognition* as an interdisciplinary hub for work on animal minds. However, cognition was here merely defined as the information processing in the brain, leaving out the subjective side of the mind. But this was not what Griffin had in mind when he criticized the remnants of the behaviourist program in the 1970s. His cognitive ethology was meant to investigate all aspects of the mind, with a special emphasis of its subjective character. He wanted to think about what the mind does *for* animals in nature. What are the problems animals solve in their natural ecological niches? As Ristau (1992), put it, the "cognitive ethologist focuses on problems faced in an animal's *natural world*; which is of course the difference between classical ethology and comparative psychology come again" (p. 125). One should therefore not mistake his call for a cognitive ethology as a mere attempt to extend the cognitive sciences to non-human animals.

Unfortunately, it took a decade after his death for such an interdisciplinary field of animal consciousness research to emerge. One hundred years after the behaviourist revolution we are finally beginning to see this chapter being written. As Birch, Schnell, and Clayton (2020) [henceforth Birch et al. 2020] put it in a recent paper in *Trends in Cognitive Sciences*: "[I]n the past 5 years, an interdisciplinary community of animal consciousness researchers, drawn from neuroscience, evolutionary biology, comparative psychology, animal welfare science, and philosophy, has begun to coalesce around" the questions of which animals have consciousness and what their subjective experiences are like? (p. 789).

The science of (human) consciousness has finally moved on from seemingly endless debates on whether or which non-human animals are conscious. We are finally rigorously studying the contents of non-human animal experience. Something of a scientific consensus has been reached that at least some non-human animals are conscious, thus opening the door for a truly Darwinian science of consciousness, such as Griffin envisioned. As Allen and Trestman (2017) put it in their Stanford Encyclopedia article on animal consciousness, we shall be forever grateful for Griffin's "crucial role in reintroducing explicit discussions of consciousness to the science of animal behavior and cognition, hence paving the way for modern investigations of the distribution and evolutionary origins of consciousness." Yet, there remain major conceptual and methodological disagreements on how we should study animal consciousness.[18] As I hope to show in this thesis, we can make progress towards a genuine science of animal consciousness by following Griffin's call to study consciousness in the spirit of an cognitive ethologist, i.e. as an adaptive phenomenon in nature.

[17] See Animal Ethics 2020 for a discussion.

[18] See Birch et al. (2022) for a recent special issue on this question.

1.4 Carrying Darwinism to Completion

Throughout the last section, my goal has been to emphasize what it means to take a naturalist approach to the place of consciousness in nature. By this I do not (only) mean the now common understanding of naturalistic thinking as the need for a strong continuity between science and philosophy, but rather the older meaning, of a *natural history* approach that begins with the careful observation of the diversity of organisms in the wild. It is in this intellectual tradition of taxonomic classifications and the mapping out of the life-histories of different organisms, rather than in laboratory experiments, that the Darwinian revolution began and changed biology forever. In order to understand organisms the natural historians began by meticulously describing the living world. This led them to appreciate the distinction between healthy and pathological variations, and eventually to a teleonomic theory to make sense of this biological normativity in a naturalistically unproblematic manner.

Hence, it ought not to be all that surprising that the natural historian Alfred Russell Wallace came up with the idea of evolution by natural selection independently from Darwin, while suffering from a fit of malarial fever on his explorations (Meyer 1895). Convinced of the common origin of species but lacking a process explanation, the fragility of his own health and vigour made him realize that varieties within populations would lead to differences in the adaptive fit of organisms to their environments and thus change species in a "struggle for existence" (Darwin and Wallace 1858, p. 54).[19] Furthermore, just like Heinroth, Lorenz, and von Frisch - Darwin himself pursued a degree in medicine. While Darwin even grew up in a family of physicians, he eventually gave up on the pursuit of medicine due to his distaste for operations and his greater obsession for natural history (Antolin 2011). Nevertheless, it is hard to believe that an appreciation for the healthy and pathological varieties of organisms did not leave an impact on his teleonomic thinking about the appearance of 'design' in nature. After all, Darwin (1859) himself maintained that we can comfort ourselves that despite the great destruction in the struggle for life, "the vigorous, the healthy, and the happy survive and multiply" (p. 29). Unfortunately, very little attention has been given to the deep link between health, evolutionary theory, and natural history outside of the evolutionary medicine movement and parts of ethology.

In order to understand healthy organisms in their natural environments, ethologists explicated the life-histories of organisms through observation and the creation of so-called 'ethograms', which were intended as an objective description of an organism's behavioural repertoire during their lifetime. Just as the taxonomists and natural historians prior to Darwin distinguished normal from pathological organisms, ethologists aimed to begin a Darwinian study of behaviour by understanding its healthy and pathological variations. It will therefore hardly be surprising that ethograms are still a common tool for both veterinarians and animal welfare scientists for detecting pathological or abnormal behaviour in animals such as tail-biting (Brunberg et al. 2011) and feather pecking (Rodenburg et al. 2003). Yet, this is not a theory of health; ethograms only constitute a list of healthy and pathological behaviours.

[19] In his biography of Wallace, Shermer (2002) offers an elegant recounting of how Wallace came to realize the phenomenon of natural selection (pp. 112-118).

It may thus not be surprising that many practitioners such as medical professionals, veterinarians, and animal welfare specialists such as Browning confidently treat health as something like a mere construct rather than a genuine integrated phenomenon in nature. They grant that we can measure parasite load, lack of nutrients, cancer, particular diseases, among other vulnerabilities and dysfunctions, and that we could rank an animal's 'health' with respect to each, but they would deny that there is some objective means of integrating them meaningfully into a single state. Something more is needed to naturalize health as a genuine whole-organism phenomenon in nature and for this we unsurprisingly require a theory of the organism.

This insight has notably already been made by Darwin's grandfather Erasmus Darwin, who was a very influential physician in a family of medical doctors, and who had an acute interest in natural history. At this time, natural history was often studied as a resource for medicine and he even anticipated some unironically 'proto-Darwinian' ideas in his treatise *Zoonomia*, a book on natural history that set out "to reduce the facts belonging to ANIMAL LIFE into classes, orders, genera, and species; and by comparing them with each other, to unravel the theory of diseases" (Darwin (1794), p. 1).[20] Such a theory was ultimately derived by Darwin, achieving his grandfathers goal to provide a "theory founded upon nature, that should bind together the scattered facts of medical knowledge, and converge into one point of view the laws of organic life" (Darwin 1794, p. 1).[21]

While Darwin himself had little to say about health and pathology, his teleonomic theory of evolution by natural selection was later used by the ethologists to likewise synthesize our scattered knowledge about normal and pathological behaviour into the very same evolutionary framework, casting their Nobel Prize for Physiology *or* Medicine in a particularly interesting light. Our folk understanding of 'health' as a biological phenomenon is revised in the light of evolutionary theory, just like that of 'design'. To advance the goal of a true biological science of consciousness, we must - similarly to the ethologists - make use of Darwin's theoretical framework to synthesize our scattered knowledge about healthy and pathological cases of phenomenological complexity across the tree of life to integrate consciousness into the final, crowning chapter of the Darwinian revolution and complete our understanding of diverse organisms, such as electrosensing platypuses, echolocating bats, and infrared-sensing snakes. Or to borrow the words of Lewontin: in order to carry Darwinism to completion we must understand the organism as both an object and subject.

Since I've repeatedly stated that state-based behavioural and life-history theory is the key theoretical resource for this task as the theory of teleonomic agency, I am now going to explain this theory in more detail and connect it to the foregoing lessons about the Darwinian revolution.

[20]Darwin denied that his grandfather's work had a major influence on him, but there are many similarities between their views.

[21]See also Nesse (2007).

1.4.1 A State-Based Behavioural and Life-History Theory of the Organism

The twenty-first century equivalent of the ethologists' attempt at building ethograms distinguishing healthy and pathological behaviours of organisms is modern state-based behavioural and life-history theory. All organisms go through a life cycle in which they are born, take in nutrients, grow, reproduce (whether sexually or asexually), and ultimately die (that is, if death hasn't already occurred before the completion of their life cycle). The diversity of life is essentially a diversity of different life-history strategies, which life-history theory aims to explain, thus making it "*the* integrative concept of organismic biology" (Kappeler 2021, p. 34). In the design-space of organisms, there is seemingly limitless room for the combinations of different traits that will lend themselves to different species-specific life-history strategies arising as optimal design-solutions to the pathological complexity organisms are faced with.

Unfortunately, philosophers of biology with the exception of Griffiths (and his recent collaborators) have given very little attention to this theory despite it being a paradigm case of adaptationist thinking and the key resource to develop a naturalist understanding of health and pathology. Inspired by a talk of Griffiths in 2018 that urged us to use life-history theory as a theory of the organism as a goal-directed system to solve the problem of distinguishing healthy from pathological traits, I aim to make progress here on the goal of developing such a teleonomic theory of the organism as an agent and an account of health as a natural phenomenon by drawing on a modern extension of this theory: state-based behavioural and life-history theory.

Life-History Theory, Agency, and Adaptationism

Originally, life-history theory was largely concerned with fairly simple models (both discrete and continuous) of the simultaneous optimization of the survival-probability at different life-stages and the number of offspring produced in each stage, in order to maximize fitness across a lifetime (Stearns 1992; Roff 1992; Morbeck et al. 1997; Roff 2002). State-dependent or state-based behavioural and life-history theory (Mangel and Clark 1986; McNamara and Houston 1996; Houston and McNamara 1999) is an important extension of this theoretical framework, since it can then be used to make the notions of health and biological normativity naturalistically unproblematic, by formalizing them in terms of fitness trade-offs in design. To understand an organism's teleonomic design is to understand their species-specific trade-offs between costly investments of resources into development, fecundity, and survival, with fitness providing an ultimate 'common currency' for this economic decision-problem, or 'game' against nature. Hence, a full understanding of an organism would be an understanding of their life-history *strategy*. As John McNamara and Houston (1996) nicely put it, life-history theory is a theory "concerned with strategic decisions over an organism's lifetime" (p. 215). It involves a naturalistically unproblematic kind of goal-directedness by treating the organisms as agents whose traits have been shaped by natural selection to contribute to a single goal of fitness-maximization:

> In life-history theory, [...] numerous aspects of an organism's life-cycle, such as the timing of reproduction or the length of its immature phase, can be

> understood by treating the organism as if it were an agent trying to maximize its expected number of offspring-or some other appropriate fitness measure-and had devised a strategy for achieving that goal.
>
> – Samir Okasha (2018, p. 10)

While adaptationism has often been criticized for atomistic thinking, life-history theory is essentially a 'holistic' kind of adaptationism in which organisms are not treated as a single adult phenotype, nor a mere robot-like bundle of traits (as Lewontin criticized the 'adaptationists') but a functionally complex and vulnerable life-cycle - a *dynamic process* that is faced with trade-offs from birth to death, for which complex optimality problems have to be solved, in short: a life-history.

Here, it is also worth responding to the objection that the view of organisms as fitness maximisers, which remains controversial both within evolutionary biology and the philosophy of biology - and especially among population geneticists who think that this is a crude simplification of the complexity of biological processes (Edwards 2007). Yet, the treatment of organisms as maximising agents is extremely widespread across many fields such as ecology and evolutionary theory and almost appears to be inevitable. Importantly, we can take on board these criticisms without denying that treating organisms as agents with the goal of fitness-maximization is a useful idealization when it comes to understanding the evolution of agency and the nature of design trade-offs. Indeed, the avoidance of such a teleonomic perspective would make it impossible to capture important normative facts about living systems that distinguish them from the rest of the non-living world.

Following Okasha (2018), we can justify this form of agential thinking that treats organisms as agents pursuing goals if they exhibit what he calls *unity of purpose*, i.e. unity "in the sense that its evolved traits contribute to a single overall goal" (p. 5). Here, we do not categorically assert that all organisms are fitness-maximizers as matter of a conteptual truism, but we simply recognize the empirical fact that evolution shaped many organisms by natural selection such that their traits work together to maximize fitness within their population. Where this is not the case and organisms are not at their local fitness-peak within their population, we can legitimately speak of these organisms as being less agent-like and less healthy from an evolutionary perspective.[22] Rather than treating these notions in a binary matter, we treat the notions of agency, health, and pathology as coming in degrees. No organism within an extant population may perfectly sit at their species-specific design optimum, but that would simply constitute the important Darwinian insight that trade-offs are inevitable and health is a question of how well an organism deals with their specific-specific pathological complexity.[23] And what is *the* teleonomic theory concerned with how organismal traits trade off against each other in order to achieve a single goal of fitness-maximization? Life-history theory.

[22]For further defenses of the organism as maximizing agent view, see Gardner (2009, 2017, 2019); Grafen (2009, 2014).

[23]Unlike the philosophy of medicine where there is often talk of some line that needs to be crossed, the Darwinian view of health and pathology rejects a threshold model. Note, however, that I am not necessarily implying that threshold thinking must be wrong when it comes to human medicine, which may serve fundamentally different purposes than the evolutionary project I am engaged in here. As I have argued elsewhere, no single concept of health may serve all these different contexts (Veit 2021c). I am here only interested in a perfectly naturalistic sense of biological normativity.

It is within this theory that we can bring out the subject-side of organisms and satisfy Lewontin's demand to bring Darwinism to completion, by paying attention to the "functional needs" of the organism (1985, p. 85). The external and internal factors come to be re-integrated with each other. As Pirotta et al. (2018) nicely put it, it is by "treating behavior as an evolutionary trait, [that] state-dependent life-history theory naturally integrates internal and external factors that are influencing individuals' decisions at multiple scales" (p. E52). Notwithstanding that adaptationist thinking *can* at times be misleading,[24] this thesis aims to show that this teleonomic theory of organismal agency provides the ideal theoretical framework in which to think about an evolutionary transition in organismal agency in the early Cambrian that made conscious evaluation *worth having*.

A Teleonomic Theory of Organisms

As I noted above, it is surprising that philosophers of biology have given very little attention to this theoretical framework, since it is here that we are provided with a teleonomic, or for that matter adaptationist, theory of the organism that enables us to distinguish biological normativity and functioning from all the other causal processes operating in living systems. No philosopher has made clearer the need for such thinking than Millikan (2002), who maintained that to understand life we need a way to distinguish the normative, functional, goal-directed processes within an organism from all the other causal processes, or mere 'noise':

> Living chunks of matter do not come, just as such, with instructions about what are allowable conditions of operation and what is to count as allowable input. Similarly, they do not come with instructions telling which changes to count as state changes within the system and which instead as damage, breakdowns or weardowns. Nor do they come with instructions about which processes either within the organism or outside it are to count as occurring within and which are irrelevant or accidental to the system.
>
> – Ruth Garrett Millikan (2002, p. 121)

To distinguish what matters to an organism, what is pathological, and what is part of it, one must attend to the organism's design. It is only with such a teleonomic theory of the organism that we can make sense of health and the subject-side of organisms. Physiologists may often succeed in progressing our understanding of the organism without evolutionary considerations, but once their work is intended to generalize, they ought to recognize that the various processes of the organism must be understood in "the light of life history theory" (Griffiths 2009, p. 23).

Morbeck et al. (1997) elegantly describe life-history theory as providing us with "a means of addressing the integration of many layers of complexity of organisms and their worlds" (p. xi). This makes it the ideal agential framework to explicate the pathological complexity thesis and make sense of the evolution of subjects and their experience. In thinking about the ecological lifestyles of different species it is thus not unreasonable to treat them as economic agents maximizing their utility (i.e. fitness). Each individual within a species is fundamentally faced with a resource allocation problem, though the solutions to this problem are admittedly more similar

[24] For critical discussions of this kind of agential thinking common to evolutionary biology, see Okasha (2018); Veit (2021a).

within a species than across species (Kappeler 2021, p. 35). This is the economy of nature.

It is no accident that ecological and economic models share many similarities and frequently borrow from each other. Under resource scarcity, there is a constant trade-off between the parameters of reproduction and survival: "reproduction in the current age class must be traded off against reproduction later; current reproduction must be also traded off against growth, and against condition (the maintenance of structures that have already developed); current growth or condition may trade off with survival to later age classes" (Griffiths and Matthewson 2018, p. 319). Boorse (1977) argued that one cannot build a theory of health from an evolutionary concept of adaptedness, since "parents hardly become healthier with each successive child, nor would anyone maintain that the healthiest traits are the ones that promote large families" (p. 548). But Griffiths and Matthewson (2018) are right to note that this is a very superficial take on evolution, insisting that there is a trade-off between offspring quality and offspring number as the British evolutionary biologist David Lack[25] demonstrated with his pioneering work on optimal clutch size (p. 305).

It is in the context of life-history theory that many puzzling philosophical questions about the nature of organisms and adaptation can be resolved, with fitness playing a crucial conceptual role for thinking about trade-offs in organismal design. By paying little attention to the actual work of evolutionary biologists, philosophers such as Boorse have made biologically uninformed statements about the viability of an evolutionary understanding of biological normativity.

Health and Pathological Complexity

Importantly, fitness is not just the number of offspring, though reproduction is correctly identified as something like the naturalized 'telos' of the organism. Griffiths and Matthewson (2018) grant that Boorse's comment may have been "light-hearted" but note that "comments such as this undoubtedly contributed to the premature rejection of evolutionary views of dysfunction" (p. 305). Indeed, they gave aid to a sphere in which a serious engagement with evolutionary theory could simply be waved away with a swift remark. But fitness is ultimately the common currency for making sense of how one organism can be healthier than another. While the health of an organism is made of of a vast range of different components, they do form a cohesive naturally existing state that can be expressed in terms of life-history theory. The mere fact that health is made up of multiple components is no reason to deny its existence as an organism-level phenomenon.

Only by employing the Darwinian notion of biological fitness are we provided with a common currency for weighing and combining the pathological complexity of different biological 'wrongs', to assess how badly (or well) things are going for an organism. The biological world is too messy for simple binary categories; be that within consciousness or health, which is why this thesis naturalizes both in terms of phenomenological and pathological *complexity*. Living systems are constantly faced with trade-offs: avoidance of one danger comes at the cost of exposing oneself to another, or entails foregoing some benefit. Health is simply a measure of how optimally an organism deals with the pathological complexity trade-offs it faces during its life cycle - or to put it differently, how well an organism succeeds in its

[25] See Lack (1947).

species-specific life-history strategy. Pathology can similar be seen as coming in degrees, i.e. a trait can be more or less pathological depending on how well it is exercising its function.[26]

Trade-offs in pathological complexity can occur on a genetic level, with selection pulling in different directions, and on a physiological level where - as Griffiths and Matthewson (2018) note - nutrients can be allocated to either somatic or germ cells. Future work on developing an evolutionary understanding of health will have to examine these and other issues in more detail, but they won't be addressed here. The focus in this thesis will be the organism as an integrated agent. Ernst Mayr (1988) once said that "the individual and not the gene must be considered the target of selection" (p. 101). Whether this is a general truth is legitimately contested, but in thinking about the evolution of evaluative agency it is certainly the right approach. The reason organisms do not start to reproduce from the instant of their birth onward, is a lack of immediate resources. Analogous to how a firm can be usefully treated as an agent, that must first gather resources, equipment, and manpower before selling on a market, organisms must first make an investment into growth before producing viable progeny. And these investments in say scales, claws, or the immune system can in turn of course have trade-offs amongst each other just like the investments of a company.

Determining how to solve the economic optimization problem between growth and reproduction under constraints was the original motivation for life-history theory. Griffiths and Matthewson (2018) point out that many organisms pursue a semelparous (as opposed to a iteroparous) strategy, i.e. "they complete all their growth before engaging in a single round of reproductive activity to which they commit all their resources" (p. 319). They mention Australian marsupials of the genus *Antechinus* as a paradigmatic example of semelparity with the death of males after a single mating season.[27] Thinking about this strange behaviour in the context of life-history theory enables us to quantify the trade-offs between survival and reproduction, and think about the teleonomic design of semelparous organisms without falling prey to human-centric thinking about what makes for 'good design'. As I point out in an application of my pathological complexity framework, while semelparous "behaviour in males may be seen as strikingly pathological, through life-history theory we can see that it is not. Their best response to their species-specific pathological complexity is to invest all their resources into reproduction in a single breeding season, and hence this not pathological" (Veit and Browning forthcoming, p. 2). Our folk understanding of health can thus be updated in the light of evolution and in particular life-history theory. Nevertheless, the *Antechinus* example is perhaps not the best illustration of an organism fine-tuned for a semelparous life-history, since the mammalian body-plan is relatively unsuited to this strategy (due to the requirement for high investment into offspring, as well as an investment into an adaptive immune system which is sometimes deliberately turned off during the breeding season) but lends itself to numerous reproductive cycles.[28]

[26]For the best recent attempt at using life history theory to distinguish healthy and pathological traits, see Griffiths and Matthewson (2018).

[27]Especially the Brown antechinus (*Antechinus stuartii*) (see Dobson 2013).

[28]However, as I have argued elsewhere, these animals provide an excellent case for how the pathological complexity framework can be used to evaluate not only species- but also sex-dependent life-history strategies, including humans: whereas female deviations from male 'norms' have often historically been labeled as 'pathological' in humans, an evolutionary understanding of their differ-

Better examples are found in insects, where many species engage in such life-cycles because a semelparous strategy relies on relatively short lives, which is favoured by a high rate of juvenile survival but a high probability of death in adulthood, with bodies being discarded relatively quickly (Fritz et al. 1982). It is because of this that insects are often claimed to not feel pain, since they would not benefit from carrying such expensive equipment for their survival (see Godfrey-Smith 2020b). This is certainly *one way* to respond to pathological complexity, though this thesis will argue against the notion that insects do not have hedonic valence. To a first approximation, however, it is true that many insects play a so-called r-strategy (as opposed to K-strategy), in which quantity as opposed to quality of offspring is maximized.[29] In the extreme, there is only one reproductive season before the organism dies, a strategy sometimes called 'big bang reproducers' or again, semelparity as opposed to iteroparity (Diamond 1982). Importantly, these should not be understood as binary distinctions, but rather as a continuum along a single axis with two extremes. Furthermore, whereas many mammals stop growing after reaching a reproductive stage, many insects, fish, and reptiles continue to grow until death (Griffiths and Matthewson 2018). These are some among many other dimensions in which self-maintenance and reproduction can be in conflict. Similar life-history trade-offs can be observed in plants, fungi, and even single-celled organisms, making pathological complexity a universal design problem that emerged at the very origins of life, as a set of teleonomic problems under constraints that need to be solved simultaneously in order to maximize fitness.

Pathological and Phenomenological Complexity

As I mentioned above, life-history models were originally fairly simple - assuming that age, reproduction, and self-maintenance can be modelled independently, i.e. that the particular differences between individuals at specific ages can be idealized away (Pianka and Parker 1975; McNamara and Houston 1996). More recent work has expanded life-history theory to incorporate behaviour, physiology, and environmental conditions of organisms, something that is now typically referred to as state-based behavioural and life-history theory (Mangel and Clark 1986; McNamara and Houston 1996; Houston and McNamara 1999). This gets us considerably closer to modeling pathological complexity as the fundamental problem of organismal trade-offs, but the problem also becomes computationally far more demanding.

To maximize fitness it is no longer just a problem of choosing a single strategy across a lifetime, but also one of choosing strategies at any moment of one's life-cycle, depending on one's bodily state and environmental conditions. The more degrees of freedom there are in the behavioural option space,[30] the higher the pathological complexity of this fitness-maximization problem, since organisms have to make sure to make the right decisions at the right time, and this depends on their current state as well as that of the environment. Indeed, they are faced with a computational

ent pathological complexity challenges can reveal such differences to be adaptive and thus healthy (Veit and Browning forthcoming). The pathological complexity thesis may thus also help us to understand differences in subjective experiences between the sexes, but because this idea is more controversial it will not be examined here.

[29] This is not to say that the r/K framework is without problems (see Nettle and Frankenhuis 2020).

[30] To put it simply: how many alternative actions an organism can take.

explosion of complexity. This complexity only increases when we add fluctuating environments such as changing weather conditions, food supply, risk of predation, and population density (e.g. Metz et al. 1992; McNamara 1997; McNamara and Houston 2008) and the frequency-dependence of optimal strategies which has been extensively studied by evolutionary game theorists (Maynard Smith 1987). These increases to pathological complexity become much harder to track, for both the organism and the biologist, but that is not to say it isn't there. Pathological complexity is the fundamental teleonomic challenge every organism has to deal with and, as I shall argue, it is precisely because of an explosion of pathological complexity during the Cambrian that sentience became a worthwhile investment, as an efficient (conscious) capacity for evaluating these continuous life-history trade-offs in action-selection. And it is this idea that I aim to explicate in Chapter 4.

Lastly, those interested in the evolution of consciousness frequently talk of a special *lifestyle* or *mode of being* - an animal way of life emerging in the Cambrian that has given rise to minimal sentience (Ginsburg and Jablonka 2019; Godfrey-Smith 2020b). Using life-history theory will help us to naturalize this idea in terms of the pathological complexity of this new lifestyle, and to assess the phenomenological complexity of animals we see around us here and now in terms of their distinct pathological complexity challenges. It is only within this evolutionary context of the optimal life-history strategies of biological agents, that we will understand organisms as objects as well as subjects, and be able to extend the Darwinian revolution towards consciousness. With this teleonomic theory of the organism in place, let us now turn our attention to the diversity of minds we find in nature.

Chapter 2

The Explanandum: Animal Consciousness and Phenomenological Complexity

> *Subjective qualities and mental experiences have remained largely untouched by the Darwinian revolution, primarily for lack of effective methods for detecting them reliably in other species, let alone analyzing them by scientific methods. But, in our present state of ignorance, we certainly cannot exclude the possibility that mental experiences, like other attributes of animals and men, exhibit continuity of variation and are not typologically discrete, all-or-nothing qualities totally restricted to a single species.*
>
> – Donald R. Griffin (1981, p. 125)

2.1 Introduction

This thesis aims to make progress in our search for the place of mind in nature by considering the evolutionary origins and distribution of consciousness across the tree of life, including the question of what the subjective experiences of other animals are *like*. The goal of this chapter is to make progress on the problem of consciousness by forcing it through the theoretical bottleneck of evolutionary theory and to explicate my notion of 'phenomenological complexity'.

If consciousness is an evolved biological phenomenon, we ought to expect both *gradations* and *varieties* of it across the tree of life - just as for any other biological phenomenon. Yet, while there is a widespread endorsement of the idea that consciousness must either be present or absent, most appear to endorse something like a threshold model of consciousness. Even philosophers of biology such as Birch (2020a) endorse the view that there is a hard dividing line in nature "between the entities that have no experiences of any kind, and those entities that do," and that "[f]inding that line, and understanding how it was crossed, is a challenge for evolutionary biology" (p. 288). But this is the very reasoning that has motivated the likes of Nagel (2012) to assert that the "Materialist Neo-Darwinian Conception of Nature is Almost Certainly False". Both are right about one thing: it is hard to find hard dividing lines within evolutionary biology. But rather than seeing this as

a challenge for evolutionary biology to explain the emergence of human-like consciousness in one fell swoop, we should consider the option that there are gradations in between that can neither be described as conscious nor non-conscious. It is of course possible that over the course of evolution there has been the accumulation of cognitive capacities that eventually led to something like a rapid 'phase transition' in which the lights suddenly 'go on', but we need to resist the view that "change on the physical biological side is smooth and gradual, and consciousness suddenly appears on top" (Godfrey-Smith 2020b, p. 264). While such events are certainly not impossible in contexts where higher-level functions are achieved through the coordination of numerous lower-level parts, thus giving rise to novel phenotypes in almost sudden manner, this is not a frequent event and we shouldn't just assume that consciousness must have emerged in such a lucky coincidence. Gradualist explanations of consciousness have inherently greater value since they are at least probabilistically more tenable.

A useful distinction could be drawn here between three different kinds of gradualist positions. Firstly, there is a weak gradualism in the form of a view that recognizes some gradations in consciousness, but firmly asserts that there is a hard dividing line between those entities that are conscious and those that aren't. Secondly, a moderate gradualism, endorsed by the likes of Birch (2020b), that recognizes differences in 'richness' across a much larger range of the phylogenetic tree including a large 'grey-zone', but remains committed to the idea that the lights are either on or off, i.e. we simply do not know where organism within this grey-zone fall. Finally, there is a strong gradualism defended by Godfrey-Smith (2020b) in his book *Metazoa* in which experience itself can be graded: consciousness itself coming gradually into existence.[1] Organisms in the grey-zone can be more or less experiential, or as Dennett (1995b) might call it: "hemi-semi-demi-pseudo-proto-quasi-minds" (p. 108). Future research on the contested cases of animal consciousness is not going to put species on either side of the alleged divide between conscious and non-conscious creatures, but rather illuminate real cases of quasi-consciousness, that should neither be categorized as conscious nor non-conscious. On pain of evolutionary continuity we ought to take seriously the possibility that there are some entities - perhaps nematodes for example - that are neither conscious nor non-consciousness, but instead better assessed as having some sort of quasi-consciousness. This is a view committed Darwinists a la Godfrey-Smith (2017a) would call a "true gradualism" since it fits the most naturally with a Darwinian picture of the world. This is the view I endorse here since it is the one that fits most naturally with an evolutionary perspective of the living world and narrows the explanatory gap by admitting many gradual steps in the evolution of the mind. Nevertheless, I will note that even if it turns out that consciousness emerged in something like a phase transition that would still be compatible with the pathological complexity thesis I defend in this thesis.

Furthermore, even if consciousness first appeared in a sudden jump, we can still maintain the value of a gradualist perspective since conscious experiences come in a great diversity of degrees and forms. How such experiential capacities became richer over evolutionary time would still require a gradualist perspective, regardless of whether a gradualist perspective of the first conscious ´sparks' is true or not. More importantly, a recognition of the phenomenological complexity we find in nature will enable us to progress in our understanding of the subjective experience of other

[1] See Veit (2022d) for my review of the book.

animals and narrow the explanatory gap by breaking the features of consciousness down into a number of different dimensions. This reshaping of the mind will not only enable us to resist the urge to think of consciousness as an all-or-nothing capacity that an evolutionary account is demanded to explain in one swift strike; but also to acknowledge that we have already developed a wide range of experimental paradigms to investigate the minds of other animals.

Chapter Outline

This chapter is structured as follows. In Section 2.2 'How to Naturalize Phenomenological Complexity?', I argue that phenomenological complexity must be at the heart of an evolutionary and comparative approach to consciousness. Sections 2.3-2.6 discuss the proposed multi-dimensional framework for animal consciousness provided by Birch et al. (2020), as well as a variety of experimental paradigms that can be used to investigate the subjectivity of other animals. This will enable us to operationalize phenomenological complexity by distinguishing five different dimensions of consciousness that can, at least in principle, all be measured in their own right. Finally, Section 2.7 'Conclusion, Objections, and Further Directions', summarizes the discussion and addresses some further difficulties in studying phenomenological complexity.

2.2 How to Naturalize Phenomenological Complexity?

In the previous chapter, I have argued that a naturalization of consciousness should not rely on conceptual analysis, but instead follow Churchland's advice to bootstrap from paradigmatic cases to develop a science of consciousness. An obvious problem with this approach, however, has been the tendency to mistake the apparent features of human consciousness as insights into necessity about all conscious experience. It is an instance of what Dennett (1991) once called the *philosopher's syndrome*: "mistaking a failure of imagination for an insight into necessity" (p. 401).

This problem is of course not unique to consciousness. As Figdor (2018) notes, "untutored imagination in general doesn't have a very good track record in terms of understanding the natural world" (p. 10). My endorsement here of an evolutionary bottom-up approach should importantly not be confused with the idea that we cannot or should not use humans as the starting point of our investigation *at all*. An insight by Lyon (2006) is important here: "Explanatory targets and starting points are not always—and perhaps not even usually—identical" (p. 52). Unfortunately, modern consciousness science has done little to distinguish human consciousness from consciousness as a natural phenomenon, suffering - as I noted in the previous chapter - from the very same problem Watson pointed out in 1913, that a focus on human consciousness would force "us into a situation similar to that which existed in biology in Darwin's time" (p. 124). However, in the face of this problem, one does not have to adopt the behaviourist response of banning consciousness entirely from science.

To study the role of consciousness in nature isn't to anthropomorphise non-human animals, it is to naturalize the mind by freeing it from the confines of the

human model of consciousness (see also Figdor 2018). Once we reject a human-centric model of subjective experience for all of consciousness, the problem of anthropomorphism will fade alongside with it. Our interest lies in what the subjective experiences of animals feel like for them, not what it would be like for us to be in an animal body. Human consciousness is a very important data-point and part of what is to be explained, but it ought not to be the be-all and end-all of a Darwinian science of animal consciousness.

Evolutionary thinking is sometimes narrowly thought of as the mere construction of hypotheses about the origins and function of traits, but it is importantly also meant to be about their diversity and the wealth of alternative life history strategies we find in nature. An evolutionary approach to subjective experience is one that recognizes both gradations and varieties of subjective experience across the tree of life, both within and across species, and studies them in a comparative manner. Consciousness, like all complex biological traits, comes in varieties and degrees so we are bound to be misled if we use human consciousness as a *model for all subjective experience* (see also Figdor 2018). Diversity in subjective experience across taxa is likely to be much higher than the diversity of subjective experience within a single species such as humans.

Yet, it may seem like an incredibly difficult task to study the variations and gradations of consciousness even if we limit ourselves to humans. Among humans there are striking differences in our sensory modalities, our emotional lives, the experienced flow of time, the ability to engage in mental time travel, how unified our experiences are, and of course pathological variations of consciousness. Note, however, that care must be taken not to characterize all variations as pathological - as 'lesser versions of consciousness' - since, as with any other biological trait, we should expect some variation to be the norm without thereby making it maladaptive.

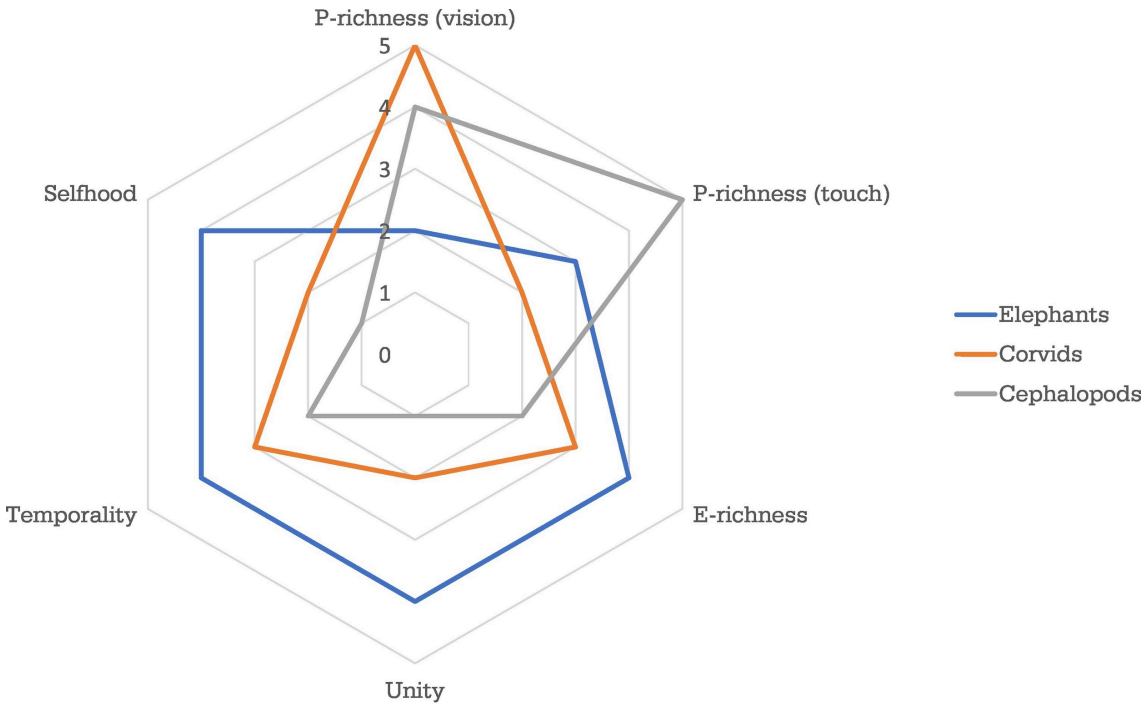
Furthermore, what counts as a pathological variation of human consciousness may not be so for an animal, e.g. what would count as colour-blindness in humans would not do so in a species such as the European mole (*Talpa europaea*) that evolved a less rich visual experience. Neither do we think that humans must be pathological because our distance vision is worse than that of a Peregrine falcon (*Falco peregrinus*). These examples illustrate that each species has their own biological norms and that we are in need for a teleonomic perspective that addresses the role consciousness plays for organisms in their normal lives, which was precisely what motivated Donald Griffin to call for a cognitive ethology. To understand consciousness in all of its diversity requires a comparative study of phenomenological complexity.

However, some may maintain that such an investigation would be impossible in non-human animals, due to their inability to verbally communicate their subjective experiences. Luckily, for purposes of this thesis, it is precisely this challenge that has motivated Birch in a recent joint paper with two scientists at the University of Cambridge - the behavioural ecologist Alexandra Schnell and Professor of comparative cognition Nicola Clayton - to offer perhaps the first attempt at assessing phenomenological complexity across the animal kingdom, by creating hypothetical 'consciousness profiles' for other animals (see Figure 2.1).

In their article "Dimensions of Animal Consciousness", Birch et al. (2020) distinguish between what they consider to be the five most important dimensions of variation in consciousness between species: *p-richness* (perceptual richness or rather

sensory richness), *e-richness* (evaluative richness), *selfhood* (self-consciousness and an awareness of other selves), *temporality* (integration of subjective experiences across time), and *unity* (integration of subjective experiences at a specific time) (p. 790). In Table 2.1, I have borrowed their list of experimental paradigms for their proposed five dimensions, on which I will elaborate in this chapter since (i) their discussion of the dimensions is quite brief, (ii) the dimensions are important for making my case that phenomenological complexity is operationalizable, and (iii) the distinctions between the five dimensions will help us in the next chapter to reverse-engineer the evolution of consciousness and narrow the explanatory gap by re-conceptualizing consciousness as a complex multidimensional phenomenon that can gradually evolve, rather than an all-or-nothing property that would appear to resist evolutionary explanation.

Figure 2.1: Birch et al.'s hypothetical consciousness profiles [reproduced from Birch et al. (2020, Figure 1, p. 791) CC BY]



"These hypothetical profiles highlight six important dimensions of variation, with p-richness represented separately for vision and touch. These are not finished, evidence-based profiles: they are conjectures based on current evidence. A key goal for animal consciousness research should be to produce a much richer evidence base for the construction of consciousness profiles and more precise ways of measuring the dimensions. Abbreviations: p-richness, perceptual richness; e-richness, evaluative richness." (Birch et al. 2020, p. 791)

2.3 The Experience of a Self

The first dimension of consciousness I will discuss here is a notion that has remained strikingly popular both in research and among the public - the idea that consciousness is self-awareness. Birch et al. (2020) call this dimension *selfhood* by which they

try to capture both an "awareness of oneself as distinct from the world outside" and an "awareness of oneself as the persisting subject of a stream of experiences, distinct from other such subjects" (p. 797). Looking at different forms of selfhood in some detail will be helpful to emphasize how even within a single dimension there can be further varieties and gradations that might require further divisions.

Theory of Mind

We begin with what is possibly the most complex form of selfhood - the awareness of other subjects with their own distinctive subjective experience - which appears extraordinarily hard to test in animals. Yet, Birch et al. (2020) are right to point to research on the so-called *theory of mind* as an important line of evidence for this capacity. This 'theory' or rather mindreading ability is meant to capture the capacity of mental state attribution to oneself and others (Baron-Cohen 1997).

The prevailing paradigm for research on mindreading has largely relied on *false-belief tasks* which test whether an animal can attribute false beliefs to others (Dennett 1987; Nichols and Stich 2003; Saxe and Kanwisher 2003). We know that humans have this ability, and much research has focused on the development of associated mental faculties during infancy (Wellman 2014; Dörrenberg et al. 2018). Furthermore, the same tests have been done in those with mental disorders that appear to have mind-reading deficits, such as autism (Baron-Cohen 2000). This focus both on development and on pathological varieties is a good starting point for understanding the role of this capacity in nature.

Unfortunately, evidence for mindreading in the animal kingdom is still sparse and seen as controversial, though there is some indication that something like a theory of mind is found in other non-human primates - especially the great apes (Premack and Woodruff 1978; Call and Tomasello 2008; Krupenye and Call 2019). The partial success in false-belief tasks may be indicative that many animals have a more rudimentary capacity to use their own experience to extrapolate to that of others, which Birch et al. (2020) refer to as *experience projection* (see Table 2.1).

Self-Awareness

When it comes to self-awareness, performance of appropriate behaviour in response to their own self-reflection in a mirror has been seen as one of the hallmark 'proofs' of animal consciousness. So it is hardly surprising that Birch et al. (2020) suggest the most sophisticated current form of this test - the *mirror-mark test* - as the key to studying this capacity (see Table 2.1). Here, an animal is tested in regard to whether it "is able to recognise a mark seen in a mirror as a mark on its own body" (Birch et al. 2020, p. 797), i.e. whether they are able to make a mental connection between the 'self in the mirror' and their own body as a self in the world.

Most human infants by around two years old are able to pass the test by touching the mark as guided by their reflection in the mirror, though there is some inconsistency in the results due varying operationalized definitions (Archer 1992; Bard et al. 2006). But this should be expected in a gradualist picture; already in healthy human development we recognize that this capacity doesn't just pop into existence, so we shouldn't be surprised if this side of experience is reported to vary across the animal kingdom. Birch et al. (2020) even highlight a recent study by Kohda et al. (2019) on the cleaner wrasse (*Labroides dimidiatus*) to suggest that "grade of

Table 2.1: Birch et al.'s suggested experimental paradigms for the five dimensions [reproduced from Birch et al. (2020, Table 1, p. 798) CC BY]

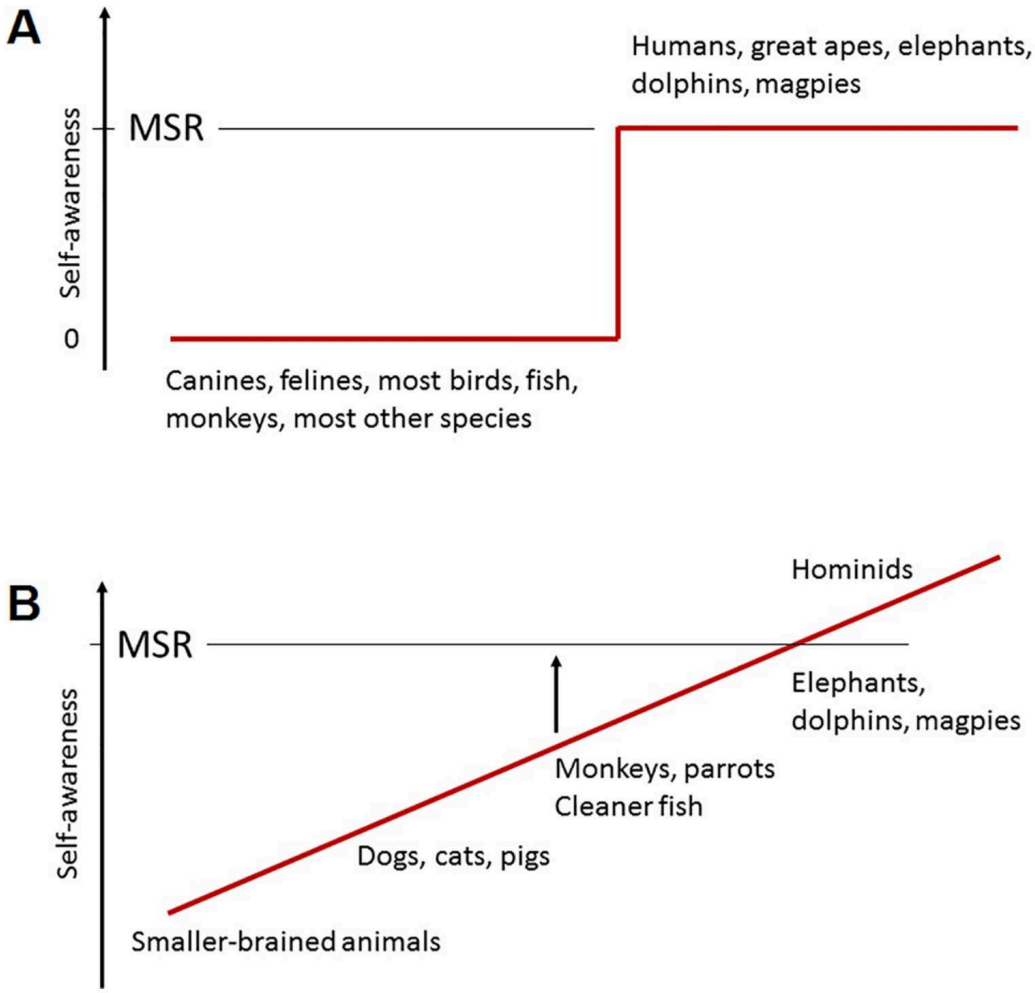
Dimension	Experimental paradigm	Question being investigated	Refs
P-richness	Induced blindsight	Can blindsight-like phenomena be induced in the animal through lesions to specific brain regions? If so, what information typically reaches those regions? (Drawback: highly invasive).	Cowey (2010)
	Discrimination learning	Can the animal learn to respond differently to very slight differences between stimuli (and how small can the differences be)?	Pearce et al. (2008)
	Reversal learning	When stimulus contingencies are reversed, can the animal rapidly learn that they have been reversed? This is potentially linked to consciousness in humans.	Bublitz et al. (2017); Travers et al. (2018)
	Trace conditioning	Can the animal still learn stimulus contingencies when the stimuli are separated by a temporal gap? This is potentially linked to consciousness in humans.	Clark et al. (2002); Allen (2017)
E-richness	Motivational trade-off	Does the animal weigh different needs against each other in a 'common currency' to make flexible decisions?	Balasko and Cabanac (1998b,a); Appel and Elwood (2009); Elwood and Appel (2009)
	Outcome devaluation and revaluation	If the value to the animal of a reward is manifestly changed, will the animal change its behaviour quickly?	Balleine and Dickinson (1998)
	Cognitive bias	Does the animal respond differently to novel stimuli depending on its affective state?	Crump et al. (2018)
	Emotional contagion	Is the animal susceptible to 'catching' the emotions of other individuals?	Osvath and Sima (2014)
Unity	Interocular transfer	If the animal is conditioned to respond to a stimulus presented in one visual hemifield, can the same response be elicited by presenting it to the other hemifield?	Ortega et al. (2008)
	Meta-control	If the two visual hemifields are presented with conflicting information, can the animal resolve the conflict?	Adam and Güntürkün (2009)
	Crossmodal integration	Can the animal integrate information from different sense modalities (e.g., vision and hearing?)	Narins et al. (2005)
	Visuo-spatial bias	Does the animal exhibit visuo-spatial biases in behaviour (e.g., a preference for using a particular eye to guide a particular task?)	Schnell et al. (2016); Rogers et al. (2013)
	Multitasking	When given two tasks simultaneously (e.g., foraging and watching for predators), does the animal divide the labour between the two hemispheres?	Rogers et al. (2013)
	Electroencephalograph studies of sleep	Does the animal exhibit unihemispheric or bihemispheric sleep?	Mascetti (2016)
Temporality (timescales <1 s)	Apparent motion	Can the animal respond differently to moving and static images? Can it make inferences from video images to real moving objects and vice versa?	Lea and Dittrich (2000)
Temporality (timescales >1 s)	Episodic-like memory	Can the animal simultaneously remember 'what', 'where', and 'when' about a specific past event?	Clayton and Dickinson (1998)
	Source memory	Can the animal remember information about how a memory was acquired (e.g., by vision or by smell)?	Billard et al. (2020)
	Memory integration	Can the animal update old memories with new information?	Clayton et al. (2001)
	Future planning	Can the animal flexibly and spontaneously plan for a future event, and for future desires, without relying on reinforcement learning?	Cheke and Clayton (2012)
Selfhood	Mirror-mark	Does the animal recognise a mark seen in a mirror as a mark on its own body?	Anderson and Gallup (2015); Morrison and Reiss (2018); Plotnik et al. (2006); Prior et al. (2008); Kohda et al. (2019)
	Body awareness	Can the animal recognise the position of its own body as a potential obstacle to success in a task?	Dale and Plotnik (2017)
	Experience projection	Can the animal predict how others are likely to behave in a scenario on the basis of a specific past experience it had in the same scenario?	Emery and Clayton (2001); Kano et al. (2019)

"A list of established experimental paradigms with the potential to provide insight into p-richness, e-richness, unity, temporality, and selfhood. There is continuing debate regarding the implications of these paradigms for questions about conscious experience. Inferences to properties of conscious states will be stronger when based on a battery of convergent experimental results from different paradigms. We restrict our attention here to established paradigms [...]" (Birch et al. 2020, p. 798)

self-consciousness required to pass the mirror-mark test is possessed by a wide range of animals" (p. 797).

What Kohda et al. (2019) demonstrated, was that the cleaner wrasse not only behaved atypically in front of the mirror, or ceased social behaviour towards the reflection, but also that they attempted to remove a coloured mark through scraping behaviour, which they did not do when the mark was translucent or in the absence of a mirror. This suggests a higher level of self-awareness than is typically assumed in fish. Naturally, there has been much controversy surrounding their experiment and the mirror-test paradigm at large (Vonk 2019; Gallup Jr and Anderson 2020) which relates to the parallel controversy over the 'correct' meaning of self-awareness (see Gallagher 2011 for a number of different definitions).[2]

Figure 2.2: A binary (**A**) vs a gradualist (**B**) view on the evolution of self-awareness [reproduced from de Waal (2019, Figure 3, p. 5) CC BY]



Much of the controversy here results from a mistaken binary conception of self-consciousness (which is partially encouraged by the very idea of the mirror *test* that can be passed or failed, as introduced by Gallup 1970). Self-consciousness, like other mental phenomena, does not just suddenly pop into existence like a light

[2]Recently, further compelling evidence for mirror self-recognition in the cleaner wrasse has been presented in a follow-up study Kohda et al. (2022), but it has not yet received critical responses.

being switched on, but gradually emerges in incremental steps towards greater cognitive and phenomenological complexity (de Waal and Ferrari 2010; de Waal 2019; Godfrey-Smith 2020c). As Frans de Waal (2019) describes it, we still "live with a ['']Big Bang['] theory, according to which this trait appeared out of the blue in just a handful of species, whereas the vast majority lacks it" (p. 1).

Borrowing de Waal's mirror-self recognition (MSR) diagram to distinguish the two views it is obvious that what he calls the "traditional binary model" (**A** in Figure 2.2) fails to recognize gradations of self-awareness. Here, self-awareness is only attributed to the Great apes, elephants, dolphins, and magpies, who are known for their high intelligence and problem-solving ability, whereas *all* others are lumped together into a single category considered to lack self-consciousness. This gross simplification of the empirical data results in much information being lost. In contrast, a gradualist view of self-consciousness recognizes a wide variety of self-recognition abilities across taxa (**B** in Figure 2.2).

All hominids appear, de Waal (2019) points out, to "spontaneously explore and play with their reflection and care about their appearance" (p. 5). Rather than looking for a distinctive kind of response that divides animals into those with self-awareness and those without, a gradualist approach would categorize a broad list of different behaviours that may be put on something like a ranking scale to assess the self-awareness of different animals. Indeed, experimental studies have already enabled us to recognize a broad range of intermediate levels of responses, where we could for instance place African Grey parrots (*Psittacus erithacus*) and monkeys (*Macaca tonkeana* and *M. fascicularis*) that have been shown to use a mirror as a tool to their advantage in order to discover objects outside of their visual field (Pepperberg et al. 1995; Anderson 1986), but do not seem to engage with themselves as a self. Even with this decades-old paradigm we are already able to recognize a diversity of more and less sophisticated notions of self-awareness:

> Reactions to mirrors range from permanent confusion about one's reflection to a certain level of understanding of how mirrors operate (e.g., using them as tools) and only brief or no confusion between one's reflection and a stranger.
>
> – Frans de Waal (2019, p. 5)

Admittedly, it will be hard to rank different responses according to how much self-awareness they exhibit. Perhaps we should, for instance, rate the discovery of objects with the aid of a mirror as richer evidence than say aggressive behaviour in response to one's own reflection. But the difficulty of these assessments should not stop us in our tracks. Indeed, I maintain that the study of self-awareness would thus be greatly improved if we'd move away from single tests towards a greater wealth of experimental set-ups to probe these gradations further. However, we should not only recognize gradations, but also further variations. As de Waal (2019) notes, these findings often involve multiple sensory modalities, since marks often involve a physical object, e.g. a sticker, rather than just a discoloration and are thus visually much more noticeable and abnormal, in addition to plausibly causing pain and discomfort, or at least some somatosensory experience.[3] This is precisely why de

[3]Another sensory modality worth considering is olfaction. Cazzolla Gatti et al. (2021), for instance, have recently used a sniff-test with wolves to demonstrate that they can engage in self-recognition based on smell.

Waal included an arrow in his depiction of the gradualist view of self-awareness (**B** in Figure 2.2), since training and multimodal stimulation are able to raise an animal towards displaying a higher level of self-awareness. A visual mark may simply not be interesting to many animals whose lifestyle naturally leads to discolorations.

For cleaner wrasse, however, visual marks are likely to be ecologically salient - the species frequently engages in cooperative social behaviour with larger host fishes, having evolved and become adapted to detecting and freeing these host fishes from parasites and dead skin. This then makes it less surprising that they are able to pass the visual mirror-mark test. Such a social lifestyle, as de Waal (2019) notes, involves complex "economic decision-making" (p. 2). Hence, within the context of my framework, an increase in phenomenological complexity in the face of this pathological complexity should not at all be surprising.

These varieties and gradations of self-awareness elegantly illustrate the importance of locating empirical findings in tests about consciousness within an evolutionary context. As de Waal (2019) puts it: "Only with a richer theory of the self and a larger test battery will we be able to determine all of the various levels of self-awareness, including where exactly fish fit in" (p. 7). To assume that a single experiment - or for that matter a set of experiments - could simply be applied across the tree of life without any consideration for the species' ecological lives and evolutionary history is as absurd as using the mirror mark test on a pig or hippopotamus (Allen 2004, p. 598). It is here that the pathological complexity thesis will offer us a useful framework for accommodating the criticism commonly raised by cognitive ethologists that the animal's ecological life-history context is being ignored within comparative psychology.

A Minimal Sense of a Bodily Self

Lastly, the most minimal form of self-consciousness Birch et al. (2020) plausibly suggest involves "registering a difference between self and other: registering some experiences as representing internal bodily events and other experiences as representing events in an external world" (p. 597). We may want to call it a bodily self. Such a capacity they suggest (similarly to de Waal) is plausibly found in all complex animals that need to be able to distinguish their own sources of bodily feedback from external ones. In addition, we may want to include features such as an awareness of oneself in space and perhaps an awareness of one's bodily state that in richer forms eventually give rise to self-awareness. Birch et al. (2020) mention *body awareness* experiments, such as those by Dale and Plotnik (2017) on elephants that try to test whether an animal is able to recognize its own body as an obstacle in a task (see Table 2.1).

Yet, determining how to distinguish conscious self-recognition from unconscious self-recognition would require a better theoretical understanding of the function of consciousness, which makes it hard to have agreement on what kinds of tests would vindicate a minimal feeling of bodily self - a problem that is admittedly shared with all the minimal forms of the five dimensions. Since we are interested in the origins of consciousness the next chapter will look again at this minimal form of selfhood and its possible connection to consciousness, but we shall now turn to the dimensions of unity.

2.4 The Unity of Experience

Birch et al. (2020) understand unity as the integration of experience *at a point in time*. By this they mean the experience of a single self or subject as opposed to multiple selves. The terminology is not perfect, however, since they use the label *temporality* for the unification of experience across time. Here, I will simply discuss them as two dimensions of unity: one I shall call synchronic, the other diachronic.

2.4.1 Synchronic Unity

We know from our own experience that healthy human adults have a highly unified experience. Smell, taste, vision, general bodily feelings - whatever it is we're experiencing appears to be experienced synchronously. Other animals, however, may not have such a unified experience.

As Birch et al. (2020) point out, the curious case of split-brain patients already suggests that the unity of consciousness may not be a necessary feature of subjective experience (see also Godfrey-Smith 2021a). In the twentieth century, some people suffering from severe forms of epilepsy had their *corpus callosum* (the connecting pathway between the two brain hemispheres) cut as an experimental treatment (Schechter 2018). Those who had it "wholly or partially severed", however, occasionally displayed strikingly "disunified behaviour when different stimuli are presented to the two halves of the visual field" (Birch et al. 2020, p. 793). Indeed, such split-brain patients appear to have two distinct streams of experience:

> If these subjects are asked to verbally describe what they see, they will report what is visible on the right-hand side of their visual field. This is because language is predominantly controlled by the brain's left hemisphere, which only has access to visual information from the right-hand side. Yet, when asked to draw with the left hand what they see, they will draw what is visible on the left-hand side of the visual field. This is because the left hand is predominantly controlled by the right hemisphere, which only has access to visual information from the left-hand side.
>
> – Birch et al. (2020, p. 793)

Researching decision-making in split-brain patients can help us to understand how a healthy brain achieves what Birch et al. (2020) call *meta-control* when faced with conflicting sensory information from the hemispheres (see Table 2.1), which may hold clues regarding the role of unity. In animals with highly lateralized brain hemispheres, however, disunified experience will not necessarily be dysfunctional, so the study of this capacity in animals whose brains resemble those of split-brain patients may reveal the adaptive benefits of synchronic disunity.

As Birch et al. (2020) point out, for animals that simultaneously engage in multiple complex tasks, i.e. what they call *multitasking* (see Table 2.1), there might be division of labour between hemispheres. A useful experimental paradigm Birch et al. (2020) mention that could probe such division of labour are *visuo-spatial bias tests* (see Table 2.1), that check whether there are preferences for using a particular sensory organ to deal with a task. Relatedly, Birch et al. (2020) suggest that *crossmodal integration* (see Table 2.1) could help to assess whether information is being integrated differently depending on the sensory modality (e.g. touch vs pain). After all, unity may not be an all or nothing affair.

What I consider to be the most important paradigm that Birch et al. (2020) discuss for this dimensions is the study of *interocular transfer* (see Table 2.1). Here, an animal is trained "to perform a task in response to a stimulus presented to one eye" after which we test "whether the task can still be performed when the stimulus is presented to the other eye" (Birch et al. 2020, p. 793). If an animal fails to do so, this could be suggestive of disunified experience, where the different brain hemispheres contain in some sense distinct 'selves'.

Finally, another compelling paradigm mentioned by Birch et al. (2020) for studying the unity of conscious experience is *electroencephalograph studies of sleep* (see Table 2.1). By studying the electrical activity of brains during sleep, we can determine whether one of the hemispheres is sleeping while the other is in a wakeful mode; i.e. whether they have bihemispheric or unihemispheric sleep. As Birch et al. (2020) rightly note, this evidence would be suggestive of "more than one stream of consciousness" (p. 794). We will return to this and other experimental paradigms in the next chapters to discuss the connection between synchronic integration and consciousness. Let us now consider the dimension of temporal unity.

2.4.2 Diachronic Unity

Birch et al. (2020) call the integration of experience across time *temporality* and illustrate it using the example of subjectively experiencing "the leaves of a tree blowing in the wind", which we do not experience as a mere "series of static snapshots" (p. 794). Beyond this integration of experience across short time-scales, however, Birch et al. (2020) also note that humans are able to engage in *mental time travel*, i.e. the ability "to recall past experiences and simulate future experiences" (p. 794). In their table of established experimental paradigms, Birch et al. divide those paradigms testing time-scales of under one second against those that integrate over a longer time span (Table 2.1). Naturally, the second category is a much more expansive one and appears to be related to memory, whereas the first is about the way we experience time. Let us thus divide them into mental time travel and the experience of time.

Mental Time Travel

Most animals are assumed to lack the ability for mental time travel - indeed, they might well be described as 'living in the moment'. So it is hardly surprising that the study of mental time travel in animals is a controversial subject and Birch et al. (2020) suspect that the ability would require "substantial cognitive sophistication" (p. 794). However, as they point out, we have already accumulated a wealth of data suggesting that corvids have this capacity, which I will discuss in chapter 5.

There are plenty of tests for the backwards-looking aspect of conscious mental time travel; or as it is typically called: 'episodic memory'. Birch et al. (2020) mention a paradigm that Clayton herself was involved in developing, i.e. testing for *episodic-like memory* (see Table 2.1), which is meant to assess whether animals are able to retrieve memory about the so-called *three Ws*: 'what', 'when', and 'where' of past events. However, the suffix '-like' was added precisely because success in such tasks does not necessarily imply conscious recall.

Birch et al. (2020) recommend that tests of *source memory* (see Table 2.1) may help us to demarcate information that can be retrieved consciously, by investigating

whether the animal can recall the sensory modality through which the information was acquired, rather than just the content. For this, the suggestion by Birch et al. (2020) to study *memory integration* (see Table 2.1) is also be helpful, since the ability to update old memories in light of new experiences could be suggestive of conscious mental time travel.

Lastly, in regards to mental time travel directed towards the future, Birch et al. (2020) suggest research on *future planning* as flexible future-directed behaviour that goes beyond mere reinforcement learning (see Table 2.1). If we can find such planning, Birch et al. (2020) are certainly right to consider it as a "promising nonverbal indicator of conscious temporal integration" (p. 796).

The Experience of Time

Could some animals perceive the world in a more fragmented way than humans? This question may seem impossible to answer. Yet, Birch et al. (2020) suggest that we could draw on research on the so-called *colour-phi illusion*, "in which two spatially separated, differently coloured dots flashed in sequence are perceived as a single moving dot that changes colour half-way across the gap", to understand how the brain integrates them to construct "a coherent account of how the stimulus is changing" (p. 794). This research supports the idea that integration of sensory stimuli plays an important functional role in humans. In some animals, however, things might look different.

Birch et al. (2020) urge us to take seriously the possibility of conducting colour-phi tests in non-human animals, even in the absence of verbal report, by suggesting that we could train animals to "respond differently to perceptions of continuous and discrete stimuli and to stimuli that change colour half-way and stimuli that do not" (p. 794). If they were then to be presented with a colour-phi test, we could change the interval between both stimuli in order to test whether - just like in the human case - there is a threshold at which animals move between an experience of the phenomena as two discrete or one continuous stimulus. Similarly, we could draw on what Birch et al. (2020) call the *apparent motion* paradigm (see Table 2.1) to assess how animals react to videos with higher and lower frame rates.

An experimental test that they surprisingly do not mention - despite already having been investigated extensively in animals and likely constituting the best evidence for assessing the experience of time in other animals - is the *critical flicker fusion frequency* paradigm, which tries to find the lowest frequency of a flickering light at which the light is perceived as continuous rather than as independent stimuli (D'Eath 1998; Healy et al. 2013; Potier et al. 2020).[4] A broad comparative study of this capacity may reveal the evolutionary benefits and drawbacks of a more fragmented way of experiencing the world, the possibility of which I will discuss in the next chapter, but we shall now turn to the dimension most readily associated with consciousness research.

[4]Schukraft (2020) offers a critical discussion of the whether critical flicker fusion frequency tracks the subjective experience of time.

2.5 Sensory Experience

The perhaps most diverse dimension of phenomenological complexity is sensory experience. Birch et al. (2020) describe the dimension of sensory experience as *perceptual richness* (or p-richness). I prefer the term sensory richness because it is less evocative of a primacy of the modality of vision, which has unfortunately so far dominated much of consciousness science. Following Birch et al. (2020), we can understand different degrees of richness in terms of how fine-grained the conscious (sensory) discriminations of an animal are.

This doesn't mean, however, that all sensory modalities can be put on a single scale. Unlike with the other dimensions, Birch et al. (2020) maintain that there are as many sub-dimensions of the sensory side of consciousness as there are different sensory modalities, such as vision, smell, and touch. Furthermore, there are sensory modalities that some animals have but humans lack (Keeley 2002). So it is perhaps unsurprising that Birch et al. (2020) deny the possibility of an overall measure for how different sensory modalities could be integrated in an overall measure: "Any measure of p-richness is specific to a sense modality, so we should not refer to a species' overall level of p-richness. A species might have richer perceptual experiences than another in one modality, but less rich experiences in a different modality" (p. 790). This is nicely illustrated in Figure 2.1 which distinguishes the hypothetical consciousness profiles of elephants, corvids, and cephalopods by splitting p-richness into vision and touch. To their credit, Birch et al. also accept that perceptual consciousness can vary within a sense modality, which (though they don't state this explicitly) could then lead to a sentience profile for a particular perceptual dimension of an animal.

Visual experience, Birch et al. (2020) note, could be divided into "bandwidth (the amount of visual content experienced at any given time), acuity (the number of just-noticeable differences to which the animal is sensitive), and categorisation power (the animal's capacity to sort perceptual properties into high-level categories)" (pp. 790-791). But their reply to the worry that this would make an overall evaluation difficult is not particularly satisfying, since they merely maintain that an animal can be considered to have a richer visual experience if it outperforms another in two but loses in one. But surely this would also depend on the case: consider a species that is mildly worse in two categories but outperforms another species in the third by a large margin. Nevertheless, Birch et al. accept that further empirical investigation may require us to move to an even more fine-grained profile or - as I would put it - a realization that there is greater phenomenological complexity in nature than we may have anticipated.

The most basic paradigm that Birch et al. (2020) mention for measuring p-richness is *discrimination learning* (see Table 2.1), which tests whether animals are able to respond to ever more fine-grained differences between different stimuli. The cognitive capacity for distinguish stimuli alone, however, may not provide us with evidence that they are also consciously experienced. To address this problem, Birch et al. (2020) suggest relying on blindsight research that has received much discussion in the consciousness literature, since those suffering from this pathological condition "report blindness in part of their visual field, but they are able to use visual information about objects in that region to guide action" (p. 791).[5]

[5]See Danckert et al. (2021) for a useful recent review.

Birch et al. (2020) argue that we could use experimentally *induced blindsight* (see Table 2.1), which has been commonly used to target the primary visual cortex (V1) of monkeys, who appear to behave strikingly similarly to humans with blindsight in *forced-choice tasks* when this part of their brain is damaged. These findings, Birch et al. (2020) argue, could be compelling evidence for sensory consciousness in animals: "if a stimulus is processed in a brain region such that damage to that region results in blindsight, then a healthy, blindsight-free animal of the species in question probably perceives that stimulus consciously" (p. 792). On pains of evolutionary continuity, this method appears quite promising. Indeed, Birch et al. suggest that we could extend this strategy "to non-mammals, based on identifying homologues or analogues of V1 in those animals" (p. 792).

Yet, they are right to admit that experimentally induced blindsight is highly invasive, difficult, and understandably considered ethically problematic despite its scientific potential. However, not all neuroscientific methods of inducing blindsight must be invasive, as recent studies on *transcranial magnetic stimulation* of the visual cortex have shown (Ro et al. 2004; Jolij and Lamme 2005; Christensen et al. 2008). While Birch et al. surprisingly do not mention this method, it may provide an excellent way to temporarily induce blindsight-like states by using magnetic impulses to disrupt the local brain regions responsible for visual experience in other animals (see Railo and Hurme 2021 for a critical review).

Another alternative method that Birch et al. propose for making progress on assessing the sensory worlds of animals is to focus on identifying special tasks that have been linked to conscious experience in humans, and apply corresponding tasks to animals. One of their proposed paradigms for this is *reversal learning* where the valence of the reward associated with two different stimuli is switched and a measure taken of the time until an animal learns the new connection, which appears to have links to conscious perception in humans (see Table 2.1).

The experimental paradigm that Birch et al. highlight for a cognitive test of sensory consciousness is *trace-conditioning*, where a conditioned and an unconditioned stimulus are presented with a short delay between them, to check whether the connection can be applied over these time-scales as opposed to merely instantaneously (see Table 2.1). As they point out, in trace conditioning tests on humans, subjects appear to only produce a conditioned response if they have been consciously aware of the initial unconditioned stimulus (such as a sound), which suggests that this form of learning is indicative of consciousness and could be applied in non-human animals. Since tests involving learning are also related to evaluation, let us now turn the last of the five dimensions.

2.6 Evaluative Experience

As I identify the origins and function of consciousness within hedonic valence, it is unsurprising that the dimension of evaluative richness (e-richness) plays a special role in this thesis, and will hence receive extra attention here. While philosophers have given relatively little attention to this dimension, several well-known scientists continue to defend the intuition of Romanes that an understanding of the evaluative mechanisms of the brain would bring us closer to understanding consciousness and suffer less from the challenge coming from the hard problem. Most prominently, the father of affective neuroscience (the neuroscience of emotions) Jaak Panksepp

(1998, 2005, 2010, 2011) argued that emotions plausibly constitute the most basic and ancient kind of consciousness - a view that will also be defended in the following chapters.

For Birch et al. (2020), this evaluative dimension of consciousness is meant to capture the subjective experience of emotions and moods - which are often described as so-called *affects* (Panksepp 2005; Browning 2020b). They suggest that 'valence' would be a great concept for understanding varieties in animal experience, since it is always involved in affective (emotional) decision-making, irrespective of speculations about whether the animal exhibits complex human-like emotions. I am less convinced of this statement since the terms 'valence' and 'affective decision-making' are used in an ambiguous way in the literature in both conscious and unconscious senses of evaluation. What they appear to have in mind is rather hedonic decision-making and hedonic valence:

> Some conscious emotions, such as pain, fear, grief, and anxiety, feel bad. These are affective experiences with negative valence. Others, such as pleasure, joy, comfort, and love, feel good. These are affective experiences with positive valence. All affective responses have positive or negative valence.
>
> – Birch et al. (2020, p. 792)

Importantly, my emphasis on adding the prefix 'hedonic' is not a mere side-note. Research in animal welfare science and the affective sciences (the sciences of emotion) has notoriously suffered from ambiguity in the use of terms such as 'fear', 'valence', 'emotion', and 'welfare', where these are sometimes used as descriptions of nonconscious neurological, behavioural, or physiological processes and at other times used to refer to subjective experiences. Prominently, this ambiguity has recently been criticized by Dawkins (2017a,b, 2021) in animal welfare science and by LeDoux (2017a,b, 2019, 2022) in affective neuroscience, both of who think that this confusion has harmed the progress of their sciences and urge caution when talking about consciousness in non-human animals.

Dawkins (2017a) has described this ambiguous use of language as "flirting with consciousness", i.e. that researchers use terms that in their everyday use carry implications of subjective experience, but when pressed in a scientific context tend to back off and state that they are not necessarily implying consciousness. Because of this, LeDoux and Dawkins believe that both the public and scientists have come to overestimate how much we really understand about conscious emotions, especially in other animals. Notably, LeDoux (2022) himself admits that he has been guilty of this way of speaking in his earlier pioneering work on how the brain achieves Pavlovian fear conditioning in rats, which has led to him routinely being introduced as the discoverer of "how conscious feelings of fear arise from the amygdala" (p. 4). LeDoux suspects that the origins of this ambiguous language lie in the behaviourists' usage of folk mental state terms (such as 'fear' conditioning) for what they saw as purely behavioural learning processes, not conscious emotions. But because the behaviourists failed to communicate their definition to the public, and more and more anti-behaviourists, such as Panksepp (2005, 2011), were interested in actually studying fear in its folk sense as a conscious experience, LeDoux (2022) asserts that we have returned to the "semantic 'wild west' of the late nineteenth century" when "animal psychology tried to use intelligent and emotional behaviour as marks of consciousness" (p. 8). As a solution, LeDoux (2022) has proposed to use terms

such as 'fear' exclusively for the subjective experience it refers to, rather than its "behavioural and physiological correlates" (p. 4). What he once called 'fear circuits', he thus now calls 'survival circuits' (LeDoux 2022).

On a first glance, the use of mental-state neutral language for what goes on in animals may appear sensible. However, in doing so LeDoux ends up artificially widening the explanatory gap between the mechanisms of the brain and the 'mysterious' subjective feelings of the mind. Admittedly, LeDoux and Dawkins are right to warn that it is very difficult to establish links between brain mechanisms and subjective experience, and that because the brain does a lot unconsciously we are in need of an explanation as to why a particular brain process is conscious rather than non-conscious. However, many, if not most, animal consciousness researchers are well aware that most information processing in the brain is done unconsciously, and that we have to disassociate conscious from unconscious processes. The fact that researchers such as Panksepp made *hypotheses* identifying subjective experiences with particular brain processes should not at all be seen as a 'Gotcha!' moment revealing the naivete of the field. The science of consciousness has made progress thus far precisely by challenging proposed hypotheses for the functions of consciousness and by showing that many proposed functions can also be performed unconsciously.

This 'speculation' is simply how science progresses, not a deep conceptual and methodological error. In a biological materialist picture of consciousness, subjective experience is constituted by particular mechanisms and processes of the nervous system. It is not 'produced' by it - such thinking would lead us straight back to dualism. But the insistence that 'flirting' with consciousness must be a bad thing and that we should not use mental state terms for physical processes inevitably brings with it older dualist ways of thinking. Similarly, the assertion by LeDoux (2019) that the problem of other minds is merely "a hypothetical philosophical argument, not a scientifically based one", and only applies to non-human animals as they have very different brains from our own, likewise evokes strangely dualist thinking about the separation of philosophy and science, and the difference between us and other animals.

LeDoux (2019) is, of course, entirely correct in stating that "if the unique aspects of our brain and our cognition are key to our kind of consciousness, then our kind of consciousness should not simply be assumed in other animals on the basis of the other minds problem" (p. 318). But his implication that the work of contemporary animal consciousness researchers is no different from nineteenth century speculations that simply anthropomorphized other animals and attributed human-like conscious experiences to them based on some behavioural similarities, is an unfair strawmanning and misrepresentation of contemporary animal minds research:

> When the claims match common sense and lore, they feel correct, and when they are repeated authoritatively in scientific or lay communities, they come to be assumed as indisputable facts. [...] The widespread assumption that innate defensive behaviours are a fool-proof reflection of conscious feelings of fear is a case in point.
>
> – Joseph LeDoux (2022, p. 8)

These assertions are greatly overstated. While members of the public might anthropomorphize their own pets, few would accept innate defensive behaviours in insects, gastropods, and worms as fool-proof evidence for consciousness. Even for more

complex evaluations and survival behaviour, few animal consciousness researchers would accept these as fool-proof indicators of conscious feelings. Indeed, there is hardly a scientific field that is kept under higher scrutiny for its hypotheses about its target phenomenon. Yet, LeDoux (2019) simply asserts that animal consciousness researchers fail to live up to the standards of human consciousness research, since their "experiments are not designed to ascertain whether a particular behaviour is consciously or non-consciously controlled. Instead, they involve amassing more and more support for the intuition that consciousness was involved" (p. 319). Worse, he accuses animal consciousness researchers of (i) not knowing the methods for distinguishing conscious from unconscious processes, (ii) not caring about them because they would support non-conscious explanations, and (iii) endorsing animal consciousness for moral rather than scientific reasons (LeDoux 2019, p. 320). I suspect that many contemporary animal consciousness researchers would take great offense at these accusations.

As this chapter illustrates, animal consciousness researchers care a great deal about disassociating conscious mental processes from unconscious ones, possessing a rich diversity of empirical approaches that put them in a very different position from the speculations of Darwin, Romanes, and co. Although future research will inevitably improve on the methods I am discussing here, they are not just the nonverbal guesswork LeDoux tries to make them out to be. That the origins of animal consciousness research have involved plenty of speculation should be seen as no more problematic than the speculations we find at the origins of any new scientific field of research, so long as they are empirically grounded and provide us with eventual tests with which to examine hypotheses.

Nevertheless, one might readily see the recent book by LeDoux (2019) as an inverse project of the one I am engaged here. Whereas I argue that consciousness arose very early in the history of animal life in order to deal with the fundamental trade-off problem of life (i.e. pathological complexity), LeDoux spends a large part of his recent book on the deep history of survival behaviours, from single-cell organisms to us, to highlight that we cannot simply infer consciousness from the presence of such behaviour and that consciousness should therefore be seen as a much more recent invention. While the survival circuits and survival behaviours of animals have become more complex over evolutionary time, he thinks that this is no reason to attribute sentience to them, since they are mere "manifestations of an ancient survival function—the ability to detect danger and respond to it" (p. 1). Here, LeDoux (2019) simply commits the same mistake as the biopsychists who use the evaluative behaviour of single-cell organisms to argue that all life is sentient. As Chapter 4 will make clear, both fail to think in evolutionary terms and recognize important gradations and transitions through which consciousness could have gradually evolved. The presence of evaluation in nonconscious organisms cannot be used as a 'proof' that *all forms* of evaluation must be nonconscious. The goal of the pathological complexity thesis is simply to defend one such hypothesis about a special kind of teleonomic complexity that made a hedonically felt form of evaluation worth having and led to an evolutionary transition in evaluative agency. While LeDoux and Dawkins may think that any such hypothesis for animals could also be explained through nonconscious means, I will argue in Chapter 4 that hedonically felt valence is the best explanation of how animals came to deal with just this kind of complexity and thus how consciousness gradually evolved.

But before I do so, let us first look at some of the empirical methods that have been developed for assessing 'valence' and 'affective states' in other animals. This is important, whether or not they are consciously experienced, since it is only by developing a greater understanding of the evaluative capacities of animals that we will be able to make sense of the evolution and function of sentience. As Birch et al. (2020) nicely put it: "Finding out how positive and negative valence are produced in an animal, and how these processes vary across taxa, should be a central goal of animal consciousness research" (p. 2972).

A particularly useful paradigm Birch et al. (2020) mention to investigate the role of affects is work on *cognitive bias* in animals, i.e. whether an animal reacts differently to new stimuli based on which affective state they are in (see Table 2.1). Relatedly, their proposal to study *outcome devaluation and revaluation* (see Table 2.1) investigates whether and how quickly an animal will change its behaviour when the rewards associated with a behaviour are changed, which is also related to the reversal learning discussed in the sensory dimension. Finding evidence for these capacities could be suggestive of evaluative experience.

Furthermore, they also list *emotional contagion*, i.e. the 'transmission' of affective states across individuals, as a promising paradigm for evaluative richness (see Table 2.1). Among highly social species that require synchronous behaviour it may even be possible to find richer evidence for this capacity than is found in humans, but this paradigm has unfortunately been understudied. Naturally, here we find a close connection to other minds research discussed in the dimension of self-consciousness.

The most significant experimental paradigm for evaluative experience, however, that Birch et al. (2020) introduce is that of *motivational trade-offs* (see Table 2.1). It is the study of how animals weigh multiple needs and opportunities against each other. Unfortunately, there has been very little comparative work on this capacity, but Birch et al. (2020) declare this paradigm a "priority for future work" (p. 793). Indeed, the importance of trade-offs is one of the core motivations for the pathological complexity thesis and will be looked at in detail in Chapter 4, but I shall lay some of the groundwork here.

Importantly, Birch et al. (2020) argue that valence provides an evaluative *common currency* for just these kinds of trade-offs in affective-decision making (p. 792). This is a different usage of the term 'common currency' from the way I have used it in Chapter 1, but they are conceptually very similar. Unlike fitness, which provides a common currency for the trade-offs of pathological complexity that organisms face in their life-histories, this usage of common currency refers to a common currency that helps organisms to evaluate trade-offs arising from their different needs/goals. There are thus two domains where common currency arguments are frequently made, one in evolutionary biology, where fitness provides an *ultimate* common currency for optimal design, and one in the behavioural and cognitive sciences, where people talk of a *mechanistic* common currency for optimal decision-making. Perhaps unsurprisingly, many authors in the behavioural, cognitive, and affective sciences have argued that complex animals have, or perhaps even must have, a *proximate* common currency linked to fitness in which the values of different actions are ranked (McFarland and Sibly 1975; McCleery 1977; McNamara and Houston 1986; Cabanac 1992; Shizgal and Conover 1996).

Again, however, we should point out that this mechanistic claim is defended in both conscious and non-conscious versions and Spurrett (2014) has offered an

elegant paper criticizing the common conflations of these common currency claims both in science and in philosophy. Indeed, the idea that hedonic valence constitutes a felt common currency of decision-making is an old one that has long been defended by philosophers in the tradition of psychological hedonism and utilitarianism, such as Bentham (1879). Here, it is also important to distinguish between the usage of pleasure as a particular kind of mental state with positive valence (as done in the above quote of Birch et al.), from pleasure as positive valence itself. Utilitarianism has often been attacked for narrowing all of human experience down into pleasure and pain, but Bentham was using them in the deliberately broad sense of hedonic valence common to all affective experience: the pleasure*s* and pain*s* of experience. This is the sense in which Chapter 4 will explicate the evolution of Benthamite creatures.

Furthermore, the idea that all affective states in humans and non-human animals can be mapped onto a two-dimensional space of valence and arousal is a widely shared view with a long tradition in the study of emotion, in both psychology and animal welfare science (Plutchik 1962, 1980; Russell and Fehr 1987; Russell and Barrett 1999; Russell 2003; Burgdorf and Panksepp 2006; Mendl et al. 2010). Rather than thinking of emotions in terms of 'typological' theories that treat them as discrete entities, such as in the work of Panksepp (2005), this tradition of 'dimensionalist' approaches treat the dimensions of valence and arousal as the real phenomena in nature, with emotions being mere theoretical constructs aimed to target particular points in these dimensional spaces.[6] Today, these ideas are especially important in Lisa Feldman Barrett's (2017) 'constructivist' theory of emotions, which had a major impact on affective psychology, in both humans and non-humans. Naturally, my approach bears greater resemblance to the tradition of the dimensionalists, rather than the typologists, since we could similarly think of conscious feelings as locations in a multi-dimensional space of phenomenological complexity.

2.7 Conclusion, Objections, and Further Directions

My goal in this chapter was to argue that any biological theory of consciousness must make sense of the full diversity and complexity of consciousness in nature, and this includes the subjective experience of non-human animals. For this purpose, I have introduced my notion of 'phenomenological complexity' as something that should be at the very heart of our theorizing about consciousness.

Against those who might be sceptical of the very viability of such a comparative project across the animal tree of life, I have reviewed and expanded the recent call by Birch et al. (2020) to develop a multi-dimensional framework for the study of animal consciousness, in order to show that phenomenological complexity can be operationalized and in principle be measured by distinguishing five dimensions of consciousness. We have finally developed a battery of tests to probe the experiences of other animals, rather than merely ask where to draw a line in the tree of life. This can be seen as great progress, even in the face of much uncertainty.

Cognitive ethologists such as Griffin were treated with much hostility, but they were right to call out the mistake in identifying a current lack of methods with the

[6] I thank Paul Griffiths for this point.

impossibility of developing them. Due to the lack of effective methods for probing the subjective experiences of non-human animals, Darwinian theorizing about the function, presence, and origin of mind has long been viewed with much suspicion. But as the present chapter has hopefully illustrated, we have reached a significant stage for a phase transition in which we can use the developed range of experimental paradigms for assessing the minds of other animals as a useful intermediate step for the removal of humans from the center of reference.

Nevertheless, a multi-dimensional framework poses new challenges, such as how many dimensions we should distinguish. An objection to the advancement of this project could certainly challenge the five dimensions discussed here. Notably, Birch et al. recognize that their dimensions, such as diachronic experience and self-consciousness, may be correlated - especially once we reach the more sophisticated levels of each - but maintain that they are conceptually distinct. Both the worry and their response are misplaced however. There is no need to have sharp dividing lines between these dimensions, as long as they offer a superior model to those offered in the past. To assert that their division constitutes genuine conceptual distinctions is beside the point. The conceptual playing field can be carved up in any number of ways, and we should be careful to not mistake such intuitively attractive distinctions for necessary insights into the nature of minds.

The mere conceivability of the possibility of animals with a "richly temporally integrated stream of experiences without any awareness of itself as the subject of those experiences" and others with "temporally fragmented 'staccato' experiences while being aware of itself as the subject of those fragment" (Birch et al. 2020, pp. 798-799) is by itself no more interesting than the assertions by Chalmers and colleagues that we could conceive of zombies that are exactly like us but lack anything it's like to be them.[7] What is needed is an evolutionary approach that asks what the adaptive value of these varying degrees and varieties of consciousness could be in order to assess their plausibility. However, I do not want to give off the impression that adaptationist thinking is the only theoretical lens offered by evolutionary biology. Most importantly, we can also build phylogenetic trees in which we sketch the plausible history of such capacities and how widely they have spread. As Calcott (2009) nicely outlines, one of the primary explanation schemes in evolutionary biology are what he calls 'lineage explanations' that try to draw plausible historical lineage scenarios for the evolution of traits. While we first need to develop better tests and a teleonomic understanding of consciousness, phylogenetic thinking will eventually enable us to make inferences about species where we have little empirical data about their cognitive capacities, yet are closely related to species where we do have plenty of evidence. Furthermore, comparative studies of such closely related species that assess their differences in physiology, ecology, behaviour, cognition, and consciousness will help us to gain further insights into which variations are related to differences in conscious experiences. Indeed, as I shall argue in Chapter 5, this feedback between animal consciousness research and biological research of species more generally is one of the main strengths of the pathological complexity thesis that will allow us to move this science further and integrate it within biology.

The five dimensions should thus primarily be seen as a useful way of overcoming current one-dimensional thinking. This is why I have not set out to fundamentally challenge their proposal. There is nothing to gain from trying to determine the right

[7] See Kirk (2021) for an excellent overview over this debate.

dimensions *now*. This must be an outcome of future comparative inquiries into the nature of mind and I expect that a greater appreciation for the phenomenological complexity in nature will force us to give up on the idea that such profiles are anything more than useful models to highlight the diversity of minds. There is no such thing as *the* "appropriate grain of analysis" (Birch et al. 2020, p. 797). Different purposes, such as which species we are comparing, will lend themselves to alternative ways of carving up the differences among them and it is these comparisons that should really matter for a cognitive ethology.

Lastly, it is precisely because of this that my review of the five dimensions began with selfhood, since this dimension most elegantly demonstrates the need for and promise of a strongly gradual and multidimensional approach to the dimensions of animal consciousness. Notably, de Waal (2019) asks us to to consider the possibility that "self-awareness develops like an onion, building layer upon layer, rather than appearing all at once" (p. 6). Some may object that such a complex nested view of a scientific phenomenon would make research much more difficult, but if consciousness is a varied and complex phenomenon in nature, we will simply have to learn to change our experimental paradigms accordingly.

Indeed, in the next chapter, I will take de Waal's metaphor of an onion to heart and argue that evaluative experience is the most promising candidate for the origins of of consciousness, by shedding the other dimensions aside one by one, in a reverse order from which I argue they plausibly emerged as something like 'outer layers' of sentience. This reverse-engineering approach will allow us to significantly reduce the explanatory gap by asking for the most minimal form of consciousness.

Chapter 3

The Origins of Consciousness or the War of the Five Dimensions

> *[I]f we contemplate the subject, we shall find it difficult or impossible to imagine a form of consciousness, however dim, which does not present, in a correspondingly undeveloped condition, the capacity of preferring some of its states to others—that is, of feeling a distinction between quiescence and vague discomfort, which, with a larger accession of the mind-element, grows into the vivid contrast between a Pleasure and a Pain. I think, therefore, it is needless to say more in justification of the level on the diagram at which I have written these words.*
>
> – George John Romanes (1883, p. 111)

3.1 Introduction

We are now in the possession of a wide range of experimental paradigms for an empirical investigation of phenomenological complexity across the animal kingdom. Furthermore, they put us in a position at which plausible sketches, hypotheses, and theories about the evolutionary origins and function of consciousness can and ought to be developed in order to transition towards a true Darwinian science of consciousness.

As the epigraph of this chapter illustrates, Romanes (1883) maintained that it would be almost impossible to conceive of the dawn of subjectivity without at least a minimal sense of evaluation, going so far as to claim that no more justification than this would be needed. While I share Romanes' view on the origins of consciousness, a modern twenty-first century account of the evolution of consciousness ought to do better than just appeal to one man's intuitions. Other possibilities need to be considered, critically evaluated, and argued against. The previous chapter gave us just these alternatives: five contenders for the crown of the most ancient kind of subjective experience. Who is the oldest in the line of succession and deserves to sit at the centre of our theory of phenomenological complexity?

By 'eliminating' one dimension after another from our rich human conception of consciousness, this chapter aims to establish hedonic evaluation as the most promising candidate for an investigation into the origins of consciousness. This Darwinian reverse-engineering approach to the problem of consciousness will allow us to turn

the human-centric methodology upside down by asking for its most humble beginnings.

Chapter Outline

This chapter is structured as follows. In Section 3.2 'Five Options for the Origins of Consciousness', I will narrow down our list of five contenders by arguing that both diachronic and synchronic unity are structural late-comer features of conscious experience unlikely to have been present at the earliest origins of subjective experience. In Section 3.3 'Down to Three', I will take on the major challenge of this chapter to make a case against both strongly internalist and strongly externalist views of consciousness that emphasize self-experience and sensory experience respectively as the most basic kinds of subjective experience. In Section 3.4 'The Last Dimension Standing: Evaluative Experience', I will defend the idea that we should turn evaluation into *the* model for consciousness as a promising teleonomic alternative to a false dilemma between internalist and externalist theories of consciousness. Moreover, I will argue that this dimension offers us a promising narrowing of the explanatory gap between matter and mind. Finally, Section 3.5 'The Spoils of War' will summarize my case for the search for the origins of consciousness in evaluation and respond to some further objections.

3.2 Five Options for the Origins of Consciousness

My treatment of consciousness as a complex multi-dimensional phenomenon is intended to narrow the explanatory gap by no longer requiring an explanation that demands all of the features of consciousness to appear together as a 'one-package deal'. A new problem, however, will inevitably arise in the sense that we are apparently faced with at least five different functions and origin stories for the evolution of experience, corresponding to each of the dimensions. Could self-consciousness, synchronic experience, diachronic experience, sensory experience, and evaluative experience all have their own independent origins? Or do some appear first, with others 'built' on top? Depending on which dimension(s) we consider to be the most basic, we may be presented with different views on the place of mind in nature - perhaps even more diverse than the results from theorizing about the function of the rich human consciousness we are all familiar with.

This is a bullet we ought to bite. In thinking about the function of consciousness, it would be a mistake to think that consciousness only does *one* thing. It has been an unfortunate development that much of the philosophical debate on functions has treated them in a binary and monist fashion (see Matthewson 2020). Traits are said to have either one function *or* another, similar to how consciousness is seen as either present or absent, without allowing for gradations and variations.[1] But traits can be shaped by numerous selection pressures and thus have a variety of functions that can be realized to various degrees.

It is therefore unsurprising that our complex human experience may not appear to be capturable with a single function statement. Indeed, it is not uncommon

[1] One excellent exception to this trend is Matthewson (2020) who argues that a trait may more or less have any single function, urging us to accept the graded nature of natural selection into our concepts of biological functions.

among naturalists such as Dennett (2005) to deny that consciousness has a function over and above the cognitive mechanisms constituting it. But such thinking already implicitly presumes a model of human consciousness and complexity. In thinking about the *role* of consciousness in nature, we must address its most humble origins. And as I shall now argue, we can at least make this problem quite a bit smaller by 'eliminating' two of the five dimensions that I grouped together under the *unity* of experience, i.e. diachronic and synchronic unity.

3.2.1 Diachronic Unity of Experience

Out of all the five dimensions, diachronic unity appears the most readily conceivable as a later feature; a layer of consciousness that may have significantly reshaped experience, but did not give rise to it. As I shall argue in this section, we should understand temporal unity as an evolved feature of the way consciousness came to be structured, not something that is necessary for experiencing itself or would somehow make consciousness 'richer'.

That diachronic unity isn't required for consciousness may seem the most obvious when we consider *mental time travel*. Unless one buys into a picture in which consciousness can *only* play a functional role if it is connected to episodic memory - or for that matter, the mental simulation of the future - this dimension appears to be strictly optional. This is not to deny that episodic memory has important functions, but that we shouldn't assert the necessity of its presence for consciousness to be functional. Here the burden of proof ought to lie with those who would want to assert that it must be present for the existence of subjective experience. After all, studies on *Severely Deficient Autobiographical Memory* (SDAM) - a condition where those with the most extreme cases completely lack the ability to *relive* their memories - show that there are humans who have subjective experience without this longer time-scale form of diachronic unity (Palombo et al. 2015; Watkins 2018). More, however, has to be said for very short time-scales of <5 seconds, or what I have called the *experience of time*.

James' (1890) description of consciousness as something like a continuous stream or flow of experience akin to a river has been influential, but it was based on human introspection, which is precisely what we seek to overcome. Admittedly, this view has empirical support. The colour-phi illusion experiments mentioned in the last chapter are suggestive that consciousness is inherently dynamic in humans. Yet, what appears like continuous movement to a human may look like a fragmented set of pictures to a different animal and what in turn looks like a series of snapshots to us may well look like continuous movement to a different species.

From a Darwinian point of view, we should thus see diachronic unity as a gradual, rather than binary matter, i.e. temporal experience can be more or less fragmented. Indeed, many animals are likely going to have a higher consciously experienced 'frame rate' than humans, which would make their experience less unified, but not inherently inferior. As Broom (2014) has once argued, we shouldn't underestimate smaller animals since "hummingbirds and mice seem to live at a much faster pace than larger, slower-moving animals such as humans" (p. 118). To support his claim, he draws on the critical flicker fusion frequency tests that I introduced in Chapter 2, which have been used by Healy et al. (2013) to demonstrate that larger animals and those with slower metabolisms have a robustly lower temporal

resolution. Naturally, a teleonomic view of mind and life should treat the subjective experience of time, the information processing resolution of time, and the speed of an organism's behaviour, as tightly linked. Smaller animals with a higher temporal resolution and faster reaction times could plausibly have the advantage of discriminating more individual frames per second than humans can, before it 'blends' into motion for them.

If animals experienced time in a more fragmented way, this might constitute an evolutionary advantage in the daily decision-problems of such animals, one that just wouldn't pay off in larger animals with slower metabolic rates and decision-making. It is entirely unclear why we should consider a blending of distinct stimuli into a temporal stream of experience as any more 'rich' or 'conscious' than the blending of my visual experience when I take my glasses off. We may even treat the identification of integration with richness as leading to an *argumentum ad absurdum*, since we would surely not want to accept that leatherback sea turtles (Dermochelys coriacea) have a richer temporal consciousness than that of humans or dogs because their flicker fusion frequency is lower and hence suggestive of a higher temporal integration.[2] Instead, a higher temporal resolution or 'fragmentation' may very well be seen as similar to higher-resolution vision. Nevertheless, there are plausible adaptive trade-offs between a higher 'frame-rate' allowing for more information being gathered per second, whereas a 'flowing', more unified mode of experiencing might enable a better understanding of causal processes. But the discussion here should cast doubt on the idea that flow-like experience must be functionally 'superior' to a more fragmented kind of experience or that such integration must have been present at the origins of consciousness.

The most reasonable take on diachronic unity would thus be to understand it as a feature of how subjective experience can be structured, something that only comes to evolve *after* consciousness has already made substantial gains in complexity. Metaphors can often be misleading, but they play a crucial role within science (Veit and Ney 2021), and we can use the common light-switch metaphor to our advantage here: the 'lights' at the dawn of animal consciousness might 'blink' on and off, only later becoming integrated into something like a stream, without thereby giving rise to an additional explanatory gap.

To accept this evolutionary reasoning is in no way to deny that the way we experience time may be crucial to what it is like to be a typical human subject and agent with long-term goals and a personal identity that is constituted by one's experiences. But I am here interested in the evolution of subjectivity and not (merely) its supposedly most rich and complex form. If we no longer require an account of the origins of minimal consciousness to contain diachronic unity, the explanatory gap has at least been partially reduced, which is all the more reason to be open to a gradualist perspective. The first sparks of subjective experience can plausibly be conceived as something more like fragmented staccato experiences without thereby making them dysfunctional. Let us thus quickly turn to the next dimension of unity.

3.2.2 Synchronic Unity of Experience

Whereas diachronic unity can be readily dismissed as a structural property that may feature in the arrangement of an organism's subjective experience, synchronic unity

[2] See Schukraft (2020).

has long been seen as one of the most fundamental features of consciousness. Indeed, many will have trouble even thinking about the idea of a disunified experience.

In our human experience, smells, tastes, images, pains, etc. simply enter into a coherent 'sphere' or 'perceptual world' of consciousness: we're living in a field, constructing an *Umwelt*. Everything that 'enters' into consciousness appears in your point of view. To even ask the question of whether there is a subjective viewpoint appears to ask for something like an integrated picture of the world, and thus presuppose synchronic unity. Again, however, we ought to be careful not to confuse an apparent feature of human experience for a necessary property of consciousness as a phenomenon in nature.

It is important to emphasize that the very terms that have established themselves for discussions about human consciousness may hinder our thinking about other animals. The 'subjective' in 'subjective experience' may be an unhelpful qualifier for minimal kinds of experience that plausibly do not come in the shape of a subjective point of view. If we consider the possibility of organisms with fragmented experiences, it becomes all too tempting to think of them as multiple little selves or subjects - a view that seems implausible and hence leads to rejecting the possibility of disunified experience. But if subjectivity is something that only gradually came into existence, this possibility should not be rejected merely from introspection into our own *subjective* experience. To further explore this claim, it will be useful to examine Giulio Tononi's *Integrated information theory* (IIT), which is the paradigmatic case of a theory that bets on synchronic integration as the core of consciousness and thus serves as a rival to my 'bet' on evaluation.

Integrated Information Theory of Consciousness

Over the last decades, IIT has been championed by Tononi and his collaborators[3] as an ontological answer to the question of what consciousness is. Drawing their inspiration from William James, Tononi and Edelman (1998) have motivated IIT by maintaining that the "fundamental aspect of the stream of consciousness" is that it is "highly unified or integrated" (p. 1846). Indeed, they go on to assert that it is impossible to conceive "of a conscious scene that is not integrated, that is, one which is not experienced from a single point of view" (Tononi and Edelman 1998, p. 1846). Instead of thinking about unity as an evolutionary 'achievement' that links diffuse phenomenological states such as a pain in one's foot and the experience of red, unity itself becomes the alleged 'fundamental property' of phenomenological experience. But is this picture plausible?

If we accept this proposed link between integration and consciousness, the following identity claim of IIT is also credible: "consciousness is one and the same thing as integrated information" (Tononi 2008), p. 232). This value is then denoted with the Greek letter for phi (Φ) as an information theoretic measure, or measurement of a system's capacity to integrate information. There are at least two ways to argue against this link between integration and consciousness: i) show that the underlying motivation is misguided, or ii) show that a theory based on this link fails to address the puzzles of consciousness. Since we are taking a naturalist approach, let us begin with the latter. However, I will note that my goal here isn't to provide

[3]See Tononi (2004, 2005, 2008, 2010, 2012a,b); Balduzzi and Tononi (2008, 2009); Tononi et al. (2016); Koch and Tononi (2011).

an exhaustive examination of IIT and its flaws. IIT is a theoretical framework that has led to a large literature including both proponents and critics, and an argument for discarding the theory would warrant at least a chapter-length treatment of its own, leaving no space to develop my own framework. Instead, I will draw on a recent target article in *Behavioral and Brain Sciences* by Merker et al. (2021) that discusses many problems of IIT, to highlight a couple of aspects of IIT that should at least decrease the confidence we may initially have in thinking about unity as a fundamental property of consciousness. But before I do so, I will first offer an abbreviated explanation of Φ.

In understanding integrated information as a measure of consciousness, Φ is defined by Tononi (2008) as "the amount of information generated by a complex of elements, above and beyond the information generated by its parts" (p. 216). However, while talk of information 'integration' and 'generation' may seem like very attractive slogans for associating IIT with the apparent unity and 'emergent' nature of consciousness, it is probably better to look at the actual mathematical construct Φ to understand what is being proposed. In what can arguably be considered the original formulation of the theory, Tononi (2004) argued that the Φ of a system has to be derived by first calculating the integrated information of *all possible* subsets made up of elements of that system. In order to explain how Φ is calculated, I will here also reuse the notation of Tononi (2004), which allows us to express the integrated information of a subset S as: $\Phi(S) = \mathrm{EI}\left({}^{\mathrm{MIB}}\mathrm{A} \leftrightharpoons \mathrm{B}\right)$. EI stands for effective information and is calculated as follows. First, we divide the elements in S into A and B. Second, we replace the states of the outputs in A with independent noise, or in information-theoretic terms: *maximum entropy* ($\mathrm{A}^{\mathrm{Hmax}}$). Third, we use the outcomes from A as inputs into B in order to assess all possible causal effects of A on B. Since A is set to be just noise, there cannot any causal effects from B on A, thus giving us $\mathrm{EI}(\mathrm{A} \rightarrow \mathrm{B})$. Fourth, we repeat these steps in reverse to derive $\mathrm{EI}(\mathrm{A} \leftarrow \mathrm{B})$. Fifth, we can add $\mathrm{EI}(\mathrm{A} \rightarrow \mathrm{B})$ and $\mathrm{EI}(\mathrm{A} \leftarrow \mathrm{B})$ together to derive $\mathrm{EI}\left(A \leftrightharpoons \mathrm{B}\right)$ for one particular sub-partition of a subset. Sixth, we repeat these steps for all other possible sub-partitions of the subset into A and B. Seventh, since an effective information value of zero would imply that we are faced with "at least two causally independent subsets, rather than with a single, integrated subset", Tononi (2004) argues that we have to find "the *minimum information bipartition* ${}^{\mathrm{MIB}}\mathrm{A} \leftrightharpoons \mathrm{B}$ of subset S" to assess the integrated information of the entire subset. However, in order to make the minimum values of different bipartitions comparable, they have to be normalized by dividing the effective information values of each bipartition by the minimum of the respective entropies: $\mathrm{H}^{\max}(\mathrm{A} \leftrightharpoons \mathrm{B}) = \min\left\{\mathrm{H}^{\max}(\mathrm{A}); \mathrm{H}^{\max}(\mathrm{B})\right\}$. Finally, the non-normalized value of this minimum information bipartition is the integrated information of the subset $\Phi(S)$.[4]

But this has only given us the Φ value of a single subset of the system. To determine Φ for the whole system, we have to repeat these steps for all other possible subsets and rank them from lowest to highest. Then, we can remove all subsets with a value of zero, as well as all those which are included in larger subsets with greater $\Phi(S)$ values since they are - as Tononi (2004) puts it - "merely parts of a larger whole". What remains are *complexes* that can integrate information and are irreducible to their components. The complex with the highest Φ value is called the "main complex" and all informational relationships going on outside of this

[4]For a more detailed explanation, see Tononi (2004).

main complex are asserted to not contribute to consciousness, which is supposed to capture the idea that "consciousness as information integration is necessarily subjective, private, and related to a single point of view or perspective" (Tononi 2004). While IIT is being continuously updated and 'suffers' from a wide range of alternative formulations, this basic idea has largely stayed the same, and I hope that this brief look at the underlying mathematics can provide the reader with a better understanding of the theory and why it has been so attractive to mathematicians and physicists. Nevertheless, it still may not seem at all intuitive as to why we should define the unity of consciousness - or for that matter, the degree/complexity of consciousness - of some integrated biological system, such as a human or a raven, in terms of this 'main complex'. However, given Tononi's background in the study of sleep, which is apparent throughout his papers on consciousness, one can perhaps understand where he is coming from. Since both consciousness and information integration fade when falling asleep and rise again during REM (rapid eye movement) sleep that can involve dreaming (Massimini et al. 2005; Tononi 2008), it does appear at least plausible to establish a link between consciousness and Φ. Furthermore, since he thinks that it is self-evidently true that all experiences are integrated wholes and cannot be reduced to their components (Tononi 2008), one can readily see why he puts forward irreducible information complexity as an identification of what consciousness *is*.

In evaluating IIT as a competitor to the pathological complexity thesis, we can readily grant that Φ as a measure of consciousness could have obvious virtues such as the *in-principle* applicability to systems very different from ourselves, allowing them to be placed along a continuum from more to less conscious (Tononi and Koch 2015). Furthermore, the idea of beginning with minimal theoretical commitments and a very simple model is familiar from other sciences studying complex phenomena. Not unlike the pathological complexity thesis, IIT may thus appear to offer a good starting point for an evolutionary approach to consciousness. Yet, while sharing some of the benefits of the pathological complexity thesis, IIT also has a set of distinct disadvantages. As I shall now argue, while information integration may be very important for consciousness, that would be a far cry from justifying the stronger claims that experience must be unified or that all of consciousness can be reduced to information integration.

Problem 1: The Axiom of Exclusion. A core problem of IIT is its overemphasis on intuitions from human consciousness. This is nicely emphasized in how they deal with the problem that an identification of consciousness with integrated information would, as Merker et al. (2021) point out, "entail the coexistence of multiple consciousnesses in a single system at a given time, rather than a single, unified consciousness, as phenomenology and parsimony considerations would seem to dictate" (p. 3). But proponents of IIT simply reject this "possibility of multiple consciousnesses in a single system by fiat", stipulating that only the main complex, i.e. the subset in the system with non-zero Φ^{MAX} (the highest Φ), is conscious (Merker et al. 2021, p. 3). Rather than taking seriously the possibility of distinct streams of experience arising from their emphasis on integrated information, it is rejected through the mere fact that human consciousness appears to be unified. Indeed, the theory hardly seems to address the problem at all, since it is at least conceivable that it will be a different subset of the brain structures or processes that at different times has the largest Φ. This would imply diachronic disunity, which would strengthen our

argument against thinking of this dimension as essential, but it also undermines the motivation by IIT proponents to justify their theory through recourse to the idea of a continuous stream of unified experiences, since the view that multiple selves take turns in being conscious does not really seem to be a defense of synchronic unity at all. This "exclusion *postulate*" leads, as Merker et al. (2022) nicely demonstrate in their work, to "some of the most metaphysically bizarre consequences of anything in consciousness theory" (p. 58).[5] Unfortunately, this is not the only instance where it seems that IIT merely postulates properties of consciousness that allegedly must be "true of every conceivable experience" (Tononi et al. 2022, p. 44), because of introspection into *our* human experience.

Problem 2: The Indicator Validation Problem. By linking Φ with consciousness, IIT appears to be able to tell us which systems are conscious and which aren't. Yet, the theory does not appear to have a means for testing that assumption. By this, I am not referring to the common criticism that Φ^{MAX} is computationally intractable and inapplicable to real brains or that the proxy measures of Φ deliver inconsistent predictions (see Merker et al. 2021). As I have argued elsewhere with Heather Browning, IIT suffers from what we called an *indicator validation problem*, since "any tests attempting to link Φ to consciousness would require prior knowledge of which systems are and are not conscious" (Browning and Veit 2020d, p. 102). As I shall show in Chapter 4, the pathological complexity thesis does not suffer from this problem since it makes a functional rather than ontological connection between complexity and consciousness, one that does not rely on the human case. It asks: what kind of complexity would make consciousness worth having? If we find systems that have high pathological complexity, yet do not have an evaluative system, that would count as striking evidence against my thesis. IIT cannot make a similar move, and instead simply asserts that any system with a non-zero Φ^{MAX} value is conscious.

Problem 3: The Bottleneck of Evolutionary Theory. Lastly, I will mention a problem with IIT that did not receive attention by Merker et al. (2021), leading me to publish the in-print debut of the pathological complexity thesis in a commentary to their target article (see Veit 2022b). In my commentary - titled "Consciousness, Complexity, and Evolution" - in reference to Tononi and Edelman's 1998 paper "Complexity and Consciousness" - I did not argue as other authors have done that IIT fails for a variety of phenomenological, formal, and neuroscientific reasons, but rather that it fails to pass the bottleneck of evolutionary theory. Admittedly, unity may well be a feature that matters in some way or another, but there is hardly any reason to think that it must be the *only* thing that matters; that they are one and the same. There appears to be an explanatory leftover, as Dennett (2019b) also notes: we must answer what consciousness does for the organism and this is a teleonomic question, not an ontological one. While Tononi (2004) states that his theory would suggest that consciousness "may have evolved precisely because it is identical with the ability to integrate a lot of information in a short period of time. If such information is about the environment, the implication is that, the more an animal is conscious, the larger the number of variables it can take into account jointly to guide its behavior", this does not in fact provide an answer to the functional question. Less integrated ways of organizing a brain may well help an organism to better guide its behaviour and IIT does not answer how consciousness itself helps the organism. A theory in which integration has, as Godfrey-Smith (2020a) nicely puts it,

[5] Since I won't repeat all their criticisms here, see Merker et al. (2021).

"no essential connection to sensing, action, and so on" (2020a, p. 209), appears biologically misguided. An evolutionary investigation of consciousness requires a sense of complexity that is linked to action, since it is here that consciousness ultimately 'pays off'. Without answering how unity helps organisms, we cannot make sense of why experience should be unified in the first place. IIT pays little attention to the most important features a biological account of consciousness should offer and thus leaves the phenomenon entirely mysterious.

Verdict: None of these problems provide a 'proof' that this theory is fundamentally on the wrong track. My goal here was only to highlight some of IIT's problems that constitute the best illustrations for why we should not just axiomatically assert that unity must be a feature of consciousness. I have argued that a theory that makes unity the fundamental property of consciousness appears to leave out some of the most important features that a theory of consciousness should have. Let us now consider some further reasons for why unity may not be all that fundamental.

Challenges to the Unity of Consciousness

In the previous chapter, I mentioned that split-brain patients are often used as an example to contemplate the possibility of de-integrated experience - with patients perhaps even constituting two conscious selves (see also Hill 2018). In a biological materialist view of consciousness we should expect differences in the organization of the nervous system and information-processing to reflect differences in consciousness. So this line of thinking should be taken seriously to avoid a fall-back into dualism, rather than be rejected from the armchair.

In asking for the function of synchronic unity, such studies of pathological cases of consciousness will doubtless be helpful and they have hinted at its role in meta-control. But if we want to learn about the possible advantages of a disunified experience, pathological cases *in humans* may be of little help, since our brain evolved to have a strikingly unified experience. Robbing it of this ability will doubtless lead to many dysfunctions, even if we can learn about surprising abilities that can still be performed with a split-brain. We could, for instance, maintain that it is artificially possible to create disunified streams of experience, while still maintaining that unity is necessary for consciousness to function well. This would no longer be an ontological unity thesis about consciousness, but a functionalist one. While the above arguments may have decreased our credence for thinking that unity must be present for consciousness, they do not necessarily undermine a teleonomic claim about the unity of consciousness. For this, it will be important to take a comparative approach and find natural experiments in the animal kingdom to study *healthy* cases of animals that could constitute natural experiments of organisms resembling split-brain patients.

Citing work by Güntürkün and Bugnyar (2016), Birch et al. suggest that we could rely on birds as an example of such natural split-brain patients since that they "have no structure akin to the corpus callosum connecting the two hemispheres of the dorsal pallium, which is homologous to the cortex in mammals" (2020, p. 793) - a structure that was long considered necessary for consciousness. In Chapter 5, I will argue that they overstate their case and that fish and non-avian reptiles make for better examples of disunified streams of experience, since they have even fewer connections between their hemispheres, but birds nevertheless constitute an excellent case for partial unity. Since birds are able to engage in meta-control and complex

goal-directed behaviours, not only despite but because of their highly lateralized brains, this raises interesting questions about how division of labour in the brain could relate to the integration of experience - perhaps providing an adaptive benefit.

Challenge 1: Interocular Transfer. Drawing on studies of interocular transfer in pigeons (*Columbia livia*) by Ortega et al. (2008), Birch et al. (2020) have argued that they may have limited integration within their vision, which can be divided into two fields: "the red field, which is the lower frontal region important for guiding pecking, and the yellow field, which covers the upper frontal and lateral regions" (pp. 793-794). As Birch et al. highlight, almost all pigeons only succeeded at having interocular transfer between the red fields, thus providing a challenge to the assumption that consciousness must be unified in other animals (see also Hill 2018). Could such partial unity provide an adaptive advantage? For birds whose eyes are placed on the sides of the head and thus provide limited visual overlap, this may not at all be unreasonable. Each eye might be engaged in a different task, e.g. predator detection vs. foraging, which may have an adaptive benefit over the single focus consciousness appears to demand in our unified experience. Excessive unity in such an animal might even be positively pathological.

Challenge 2: Unihemispheric Sleep. Recall the electroencephalograph studies of sleep from Chapter 2, which have demonstrated that some animals let one hemisphere sleep, while the other stays awake. Unlike humans, for instance, aquatic mammals like seals and dolphins appear to have unihemispheric sleep (Mascetti 2016). Birch et al. (2020) consider this to be further evidence challenging the necessity of unity, since birds have also been shown to engage in unihemispheric sleep. If it is possible to have two streams of consciousness, this could well entail the possibility of dream experiences in one brain, while the other is awake and consciously evaluating things going on around the animal. Obviously, more research will have to be undertaken to answer such questions, but given the presence of division of labour in brains we cannot simply assume that consciousness must be a unified experience in order to be functional. Rather than being necessarily pathological, disunified streams of experience may well offer their own adaptive benefits without thereby being 'less' conscious. Notably, recent research has suggested that great frigatebirds (*Fregata minor*) are able to sleep in this unihemispheric manner even during flight (Rattenborg et al. 2016). Since flight and swimming contain some similarities, such as the ability to move uninterruptedly in three dimensions for long distances, there might be a case to be made for convergent evolution of unihemispheric sleep in seals, dolphins, and birds. There is a pathological complexity trade-off between rest and attention that different species appear to solve in different ways.

Challenge 3: Division of Labour and Synesthesia. Lastly, it is worth noting that despite the fact that research on meta-control, crossmodal integration, visuospatial bias, and multitasking in animals with lateralized brains may not directly challenge the alleged unity of consciousness, this research can reveal the adaptive benefits of having a brain that is not highly integrated, such as division of labour, and thereby at least indirectly challenge the assumption that unity must be fundamental to consciousness. The most striking cases of lateralization, after all, are found in vertebrates, i.e. in fish and reptiles, that evolved from animals with more 'unified' brains. And the above research should at least make us take seriously the possibility that this division of labour is associated with disunified experiencing. In order to think about the experience of such natural split-brains, Hill (2018) offers

a useful metaphor of *islands* representing a kind of clustering of particular experiences, "where the components of an island are unified with each other but not with the components of other islands" (p. 4). This metaphor is a good one because it can help us to recognize that even in humans, our islands of experiences may not be all that perfectly integrated. While our visual field very well may be - after all this is the sensory dimension that gave rise to the focus on unity - my experience of sounds, smells, and bodily feelings also appear to be strikingly distinct. They do not perfectly melt together as they might do in someone with *synesthesia*, i.e. subjects who can smell sounds, see shapes for smells, have colours associated with words, and so on (Ward 2013). It seems hard to deny that the subjective experience of someone with synesthesia shows even higher unification, casting further doubt on the idea that unity is a binary, suggesting instead it may come in varieties and degrees. Upon reflection, I can perfectly grant that there is a sense in which I have multiple selves - one with a visual point of view, one with goals and motivations, another with a sense of bodily self - that only partially integrate with each other into something like a unified self. That does not appear to be a conceptual error and is quite reasonably necessary for these capacities to play their functional roles. Instead of thinking of human experiences as a wholly unified sphere, it might be better to think of them akin to the partial unity exhibited in the interocular transfer experiments with the visual fields of pigeons. Tentative evidence suggests that synesthetes have improved learning and memory (Watson et al. 2014), but also that some suffer from sensory overload and the interference of synesthetic experiences with tasks such as for example mathematics (Rich et al. 2005), which could help us to understand the trade-offs in having more or less unified experience. Whereas some severe forms of synesthesia are probably pathological, others are plausibly part of normal healthy human variation. Studying these phenomena in more detail will help us to better understand the benefits and drawbacks of unity, and put us in a better position to think about how much unity to expect in animals with very different kinds of ecological lifestyles. We should not pretend that unity must be fundamental prior to such empirical investigations.

3.2.3 Unity and Consciousness

Ultimately, this section hoped to demonstrate that the apparent unity of consciousness in humans, whether of the diachronic or synchronic kind, cannot be treated as an insight into necessity regarding what consciousness must be like. Previously, I mentioned that earlier discussions in the philosophy of mind distinguished between 'consciousness' and 'qualia' as two distinct problems. It might be worth returning to this older way of thinking, a way that didn't treat the rich properties of human-like consciousness as necessary insights into qualia, which might be much more rudimentary, and in principle present in animals very evolutionarily distant from us (see Godfrey-Smith 2016c,a). Merging these into a single problem may not have constituted progress after all, with the existence of partial unity challenging preconceived notions that consciousness necessarily belongs to discrete subjects. This is why I find the term subjective experience less helpful than have others such as Godfrey-Smith in thinking about the origins and most minimal forms of qualitative experience.[6]

Rather than treating 'qualia' as a concept that should be analyzed through

[6]See Godfrey-Smith (2021a) for a discussion of these issues.

introspection into human consciousness - and thus perhaps inevitably merging these two phenomena together before even reaching consensus on how qualia should be conceptualized (see Tye 2021 for a review) - my goal is to naturalize this notion through an investigation into the most rudimentary beginnings on the evolutionary path towards the complexities of human consciousness. In the eyes of philosophers sitting in the Nagel and Chalmers camp this might make me an eliminativist, but a better description of this approach would be revisionist, since we revise the notion of qualia in a naturalistically unproblematic manner. As unity does not appear to be necessary for consciousness, we can narrow the explanatory gap by shelving it off from our investigation of the very origins of consciousness.

Indeed, when we consider the most simple forms of subjective experience, unity appears to be in a sense a more trivial affair of definition, rather than an interesting biological property regarding how experience comes to be organized. And these two senses need to be distinguished. If an animal's subjective experiences are limited to a distinction between two senses (e.g. hot and cold, or evaluations of good and bad) there is little point speaking of unity or disunity, simply because there are so few distinct subjective experiences to begin with. Such an animal may have a Φ value high enough to warrant the attribution of consciousness, but it is not unity that does the explanatory work here, since it would be fundamentally in sensing or evaluation that consciousness pays off for the organism. The fact that nervous systems have *some degree* of unity or integration that gives rise to Φ, has little to do with consciousness, and everything with the fact that all life must show unity at least to some extent. The mere fact that there is "correlation between levels of conscious arousal and phi" (Ginsburg and Jablonka 2019, p. 126), should therefore not be seen as providing significant support for IIT's insistence that unity is the fundamental property of consciousness.

A passage in Godfrey-Smith (2016d) expresses this quite nicely: "To some degree, unity is inevitable in a living agent: an animal is a whole, a physical object keeping itself alive. But in other ways, unity is optional, an achievement, an invention" (p. 87). And it is these latter, optional, forms of unity that matter for thinking about whether consciousness *must* be unified. In these functional senses of unity, it becomes reasonable to think that the unity of "human experiential profiles might be a later-arriving, derived trait" (2020a, p. 209). We can grant unity may well be a very important feature of human consciousness - perhaps even something that would almost inevitably evolve in systems with heterogeneous washes of qualia - but we should not expect it as a fundamental functional feature of consciousness that explains its raison d'être. From a teleonomic perspective, both dimensions of unity remain of little functional importance until there are animals with a certain degree of phenomenological complexity that allows subjective experience to be organized in interesting new ways to play particular functional roles. Until then, one can hardly speak of integrated organization at all, any more than we should praise an 8 year-old for having a diversified financial portfolio with their two five-dollar notes stored in different places in their room.

Both diachronic unity and synchronic unity appear to be features of the way experiences are organized, not something we need to explain if we are merely concerned about the first sparks of experience in their own right. They do not appear to solve the hard problem of why there are some states that *feel* like something to an organism. While those betting on unity as the fundamental dimension to

solve the problems of consciousness may well turn out to be right, it appears more plausible from an evolutionary point of view to treat forms of unity as later 'add-ons'. Tononi et al. (2022) allege that unity must be seen as a fundamental property of consciousness in any "satisfactory explanation of consciousness" (p. 44), but a satisfactory account of unity can explain it as a special arrangement of the way in which consciousness is organized in humans, without thereby making it a fundamental property of consciousness. Again, we need to be careful not to fall prey to the all-too-common confusion between human consciousness and consciousness as a phenomenon in nature. And since an 'elimination' of unity from the necessary properties of consciousness substantially narrows the explanatory gap, this alone would warrant the search for a theory of consciousness in more basic experiences.

3.3 Down to Three

This leaves us with three remaining dimensions of consciousness: i) the experience of a self, ii) sensory experience, and iii) evaluative experience. To think that these must all come together as a one-package deal for consciousness, would still leave the explanatory gap incredibly wide. In thinking about the origins of minimal consciousness it is hence plausible to think that it is in only one of these dimensions that consciousness arose.

Despite the fact that each dimension can provide us with a distinctive model of consciousness, however, the literature has largely focused on the first two options. As I shall argue, this has placed the science of consciousness in an unfortunate dilemma between internalist and externalist approaches to consciousness, i.e. that consciousness is to be explained either through recourse of properties internal to the organism or external to it. While these theories may well succeed at explaining parts of the phenomenon, I will make a case that there remains an explanatory leftover, i.e. precisely what the other dimensions were meant to explain, and thus giving off the impression of an explanatory gap.

My arguments in this section are not intended to rule out or eliminate theories of consciousness focusing on selfhood or sensory experience, but rather to undermine an almost certain confidence seen in parts of the field that it is in these dimensions that the problems of consciousness are resolved. This will help us to motivate looking at an alternative dimension that has the benefits of each dimension, while not sharing their respective weaknesses: the dimension of evaluation.

3.3.1 Experience of a Self

Consciousness was once seen as a higher-order form of thought or awareness of oneself as an individual. Moreover, it was treated as something that should distinguish humans from other animals and could perhaps be found in some rudimentary fashion in our close relatives, such as chimpanzees and bonobos, or other intelligent animals such as dolphins and elephants able to recognize themselves in mirrors (Keenan et al. 2003). However, this kind of experience is now more typically regarded as a special kind of subjective experience that is only present in some sentient creatures, the experience of self-consciousness. Indeed, the historical association of the term 'consciousness' with some form of rich and exclusively human experience, is precisely

why many of the researchers engaged in the study of animal consciousness prefer the term 'sentience' to refer to more basic kinds of experiences.

Yet, it would be easy to dismiss the dimension of self-hood by appealing to the radical demands of the richest kinds found within its bounds and to contrast these with a simple evaluative feeling. Just as with the dimensions of unity, we should consider the most rudimentary forms within this dimension, i.e. a minimal sense of a bodily self, rather than comparatively rich self-awareness. Here, it will be useful to examine a direct competitor to the pathological complexity thesis that can be found in the so-called *autopoietic* tradition.

Autopoiesis and Consciousness

The autopoietic tradition originated with Humberto Maturana and Francisco Varela, who maintained that all "living systems are cognitive systems and living as a process is a process of cognition" (1980, p. 13). A continuity thesis between mind and life can hardly be stronger than this, so it will not be surprising that this tradition has also tried to address the problem of consciousness, perhaps most influentially by Evan Thompson (2007) in his book *Mind in Life: Biology, Phenomenology, and the Sciences of the Mind*.

Maturana and Varela (1980) introduced the term 'autopoiesis', which translates as 'self-creation', and was intended to capture the organismal phenomena of 'viability', 'self-maintenance', 'self-organization', 'self-production', and 'self-reproduction'. It is these phenomena that they saw being neglected in an alleged object-like view from the Darwinian perspective that neglected the autonomy and subject-like nature of biological organisms. Life, the autopoietic tradition emphasizes, must be bounded and self-maintaining in order to allow its continuing existence and these features should accordingly be the centre of our attention.

Such a view of life emphasizing autonomy and subjecthood certainly lends itself to thinking of life as a bridge between matter and mind. Thompson maintains that consciousness can be described as "a kind of primitively self-aware liveliness or animation of the body" (p. 161), which arises from the "autopoietic identity and sense-making of living beings, but in addition it implies a feeling of self and world" (p. 221). Organisms are seen as being engaged in the "self-production of an inside that also specifies an outside to which it is normatively related" (Thompson 2007, p. 163). Instead of being passive objects subject to external forces, here they are seen as active subjects determining their own fates with a high degree of *autonomy*. One can readily see why such a framework would lend itself to thinking about the origins of *subjective* experience and why Thompson (2007) frequently invokes Lewontin to argue that the Darwinian/functionalist view of life and mind is mistaken, or at least needs to be supplanted. This underlying motivation of the autopoietic tradition, to provide an internalist alternative to the perceived externalism of Darwinism, is also reflected in related attempts of the tradition to make sense of health, agency, and biological normativity, providing us with a challenge to the pathological complexity thesis on multiple issues of interest.

Granted, it is tempting to base a theory of the organism on these normative notions of vital self-organization. No longer do naturalistic explications of this idea suffer from a failure to distinguish such thermodynamically open systems able to maintain themselves, from other systems such as convection cells, hurricanes, flames, and the like. While these also engage in an exchange of energy and matter to main-

tain themselves away from thermodynamic equilibrium, Nicholson (2014) rightly notes that organisms "are distinctive in that they regulate their interaction with the outside environment by means of physical boundaries (like membranes or skin) that they themselves generate" (p. 355). It is perhaps unsurprising that self-organization accounts have been met with much enthusiasm, with many thinking that they "hold the key to naturalizing rather elusive notions like function, normativity, and agency" (Nicholson 2014, p. 355). I would happily concur *if* self-organization was the goal of the organism. For systems designed to achieve self-organization as their raison d'être, this would be the best way to naturalize these notions. But living systems have a different rationale.

Problem 1: A Mistaken View of Life. Like the pathological complexity thesis, the work of Thompson (2007) is an attempt to solve the problems of mind within an account of life, ultimately trying to use the properties of life as something like a conceptual bridge to narrow the explanatory gap between matter and mind. But whereas the pathological complexity thesis relies on a Darwinian life-history view of living systems, the autopoietic attempts to provide an alternative to a Darwinian view of life. The problem for such approaches, however, is that the goal of biological systems is ultimately reproduction.

For this, one does not have to deny that 'self-maintenance', 'self-organization', and 'self-production' are real phenomena in biological systems. But they are likewise processes that serve the underlying purpose of the organism, i.e. to participate in a 'game of fitness maximization'. An adaptationist does not have to demur to the ontological points popular in this literature which claim that "organisms are more accurately conceived as dynamic processes than as stable things" or that the "identity of organisms hinges on the fact that they are continuously undergoing change" due to the fact that they have to deal with thermodynamics (Nicholson 2014, p. 355). Just because many evolutionary models idealize or "abstract away the temporal dimension of organisms so that they can be studied as static things", does not mean, as Nicholson (2014) seems to suggest, that evolutionary biologists fail to recognize that such idealizations "do not come without costs" (p. 356). This is simply not true. Modern evolutionary theory no longer perceives organisms as mere survival machines. Life cycles are processes, and life history is a central theory of evolutionary biology upon which to build a naturalist Darwinian theory of the organism. As Griffiths (2009) once argued in a paper with the title "In What Sense Does 'Nothing Make Sense Except in the Light of Evolution'?", it is only within the context of evolution by natural selection that we truly understand the teleonomic nature of organisms:

> Non-evolutionary accounts of biological functioning draw the boundary between biological functioning and other, irrelevant causal processes in which organisms are participants in the wrong place. They exclude activities which no-one can seriously doubt are examples of biological functioning. This happens because 'viability', 'self-reproduction' and the rest are only one component of a more encompassing ability which involves activities whose focus is not on the maintenance of the physiological individual.
>
> – Paul Griffiths (2009, p. 22)

This quote nicely recapitulates the insights by Millikan that were discussed in Chapter 1: that lots of biological goings-on of organisms are not concerned with self-

maintenance but maximizing the number of viable offspring. To this end, the organism will take the risk of undermining its own bodily systems and perhaps even hastening death. One cannot understand the organism - or for that matter biological agency, normativity, and function - without accounting for these Darwinian design trade-offs, or one will positively misunderstand what parts and processes of the organism are *for* and when something goes *wrong*.

Without the Darwinian point of view we would inevitably end up mischaracterizing the various trade-offs going on in biological design precisely for the real purpose of the organism, which is to maximize its representation in future populations. As Griffiths (2009) once forcefully argued, the mechanisms and processes underlying 'big bang mating' strategies "obviously do not contribute to the capacity of individual males to maintain their form. But they do contribute to the life-history strategy by which these males maximize their contributions to future generations" (p. 22). Self-maintenance up until that point is merely a means to an end. If one is interested in the goal-directedness of living systems, one has to understand that it is the organism as an integrated whole that has been shaped by natural selection to contribute to a single goal of fitness-maximization, not self-organization.

Problem 2: A Mistaken View of Darwinism. The emphasis on life being inherently dynamic, goal-directed, and normative is something similarly emphasized here as a useful path to reshaping the materialist side of the mind-matter problem. But the autopoietic tradition goes too far in various respects, over-reaching in almost the opposite direction from that of the 'mainstream' in the philosophy of mind, a fact which owes itself to its origins and conception as a 'radical' challenge and alternative to the Darwinian view of life (Escobar 2012).

It is thus perhaps not surprising that Thompson (2007) explicitly contrasts the autopoietic strategy as an alternative to the Darwinian view of life. In a reply to his commentators, he contrasts the Darwinian reverse-engineering perspective with an autonomy perspective (Thompson 2011, p. 177). Thompson asserts that the evolutionist's method of reverse engineering is a mere interpretive framework or heuristic and that Darwinians make the mistake of extrapolating from its success as a method to the claim that organisms are design artifacts (2007, p. 211). Instead, he asserts the *autonomy* of life as the fundamental property missed by the functionalist approach. This may be tempting, since the functionalist approach is derided by many as insufficient to account for the basic problem of experience. But the 'autonomy-approach', it turns out, is simply empirically inadequate. Owing to its emphasis on an internalist alternative to what are perceived as excessively externalist ways of thinking about life and mind, the approach misses both the fundamentally teleonomic nature of organisms and the reciprocal influence between organism and the world.

Thompson's argument is reminiscent of older vitalist debates and criticisms of evolutionary-functionalist approaches to consciousness for supposedly being unable to capture the most fundamental property of life and mind. It is merely asserted that autonomy is fundamental and sufficient to explain both life and mind. But as I've argued in Chapter 1, the very motivation of conceiving Darwinism (or for that matter, adaptationism) as a necessarily externalist approach that requires a radical alternative, is simply a failure to distinguish the externalist pre-Darwinian design thinking with the more recent teleonomic thinking about design and normativity that conceptually re-engineered them in terms of natural selection and feedback

between organism and environment.[7]

Problem 3: A Resistance to Adaptationist Thinking. A life-mind continuity thesis could hardly be stronger than to identify life with a process of cognition. From here, it is only a small step towards a full embrace of biopsychism and to grant consciousness to all life. This would be an odd evolutionary journey to say the least, but it is being taken seriously by authors such as Thompson (2022). We take one step in the emergence of life and are suddenly presented with consciousness.

Like Godfrey-Smith (2020b), I fear that this tradition makes consciousness too much of "an automatic feature of just being a living organism located in the world" (p. 117), even if its attribution to the most minimal living systems is avoided. Consciousness in such a picture comes in a certain sense for free, which may be intuitively more attractive than a gradualist model, but shares some uncomfortable parallels with epiphenomenalist ideas. It does not seem to give us any purchase on the raison d'être of consciousness. We are not given a gradualist story of how consciousness (or for that matter, agency) emerges, rather it is something that simply comes *along* with or rather *constitutes* a certain kind of living activity. This makes it hard to think about things gradually becoming more agential, experiential, or - to use their language - 'autonomous', precisely because of the resistance to adaptationist thinking.

Indeed, it is hard to make sense of the idea that a vague wash of feeling of 'presence in the world' would become refined and enriched through the process of natural selection if these experiences represent neither useful sensory nor evaluative information. The mistaken idea of avoiding the hard problem by trying to rely on a *non-functionalist explanation* makes it ultimately impossible to explain the evolved gradations and variations of subjective experience across the tree of life and this is why we should at least be skeptical that a search for the origins of consciousness in this dimension will succeed. Even if the approach could make sense of the qualitative experience of a self, it is unclear how the other dimensions can be built on a model of self-experience without - again - making them mere automatic features of living activity. We do not appear to be able to answer the question of why some sensory and evaluative processes are felt whereas others are not. The resistance to adaptationist thinking makes us unable to think about the most important properties of consciousness, i.e. what it does *for* organisms, and thus deprives us of a unique explanatory strategy for the life sciences.

Problem 4: A Neglect of Feedback. Lastly, in trying to motivate a view of life as active and autonomous agency, external goings-on come to be endogenized as just another kind of organismal activity, leading to a neglect of the "to-and-fro traffic characteristic of organism/environment relations" (Godfrey-Smith 2016b, p. 778). Instead of making the externalist move of seeing internal events and processes as just another kind of environment, they neglect the role of the organism's environment in an effort to not undermine the 'autonomy' of the organism. The environment is almost seen as just another kind of internal goings-on, the organism forcing itself onto nature, without reciprocal influence. Thompson (2007) urges us to see autopoietic systems as "sources of their own activity, specifying their own domains of interaction, not as transducers or functions for converting input instructions into output

[7] Tanaka et al. (2020), for instance, show that Lewontin's arguments for the role of the organism as a subject can be readily assimilated into more complex models of adaptive landscapes that include feedback between organism and environment.

products" (p. 46). A strikingly internalist picture is provided in which organismal properties are said to emerge between top-down and bottom-up processes within the organism, in a manner of supposedly naturalistically unproblematic 'circular causality' that has been seen with suspicion by more reductionist-inclined scientists.

Godfrey-Smith (2016b) describes writers in this family of views as being overly zealous in their determination to describe and explain living systems without any reference to passivity. In an anti-externalist mantra, they argue that organisms must be seen as subjects *instead* of objects, neglecting the possibility that agency can come in degrees, and that organisms are both subjects *and* objects. Indeed, Thompson (2007) tellingly describes the nature of any autopoietic system as being "defined by its endogenous, self-organizing and self-controlling dynamics" which do "not have inputs and outputs in the usual sense" (p. 43). While recent work on the mechanisms of such self-organizing systems do not deny the interaction between organism and environment - highlighting dynamic feedback, and the need of such systems to maintain organizational and operational and organizational closure in order to stave off thermodynamic processes towards entropy - the emphasis on autonomy has led, as Godfrey-Smith (2016b) notes, to a continuing resistance "to the role of ecology, in a broad sense – resistant to the fact that it is part of the nature of life to be in ongoing interaction with an environment that is *other*" (p. 778). But this neglect of ecological feedback is precisely why the theory appears unable to explain the heterogeneity and adaptive fit of subjective experience to the life-histories of animals.

Verdict: Again, none of these problems must spell doom for the autopoietic approach. While their opposition to Darwinism characterised as a necessarily externalist program is mistaken, that doesn't necessarily mean that their internalist approach to the mind is. Nevertheless, the resistance of the approach to adaptationist and functionalist thinking makes it hard to see how they could account for the dimensions of sensory and evaluative experience. If we are interested in a strong continuity thesis between life and mind, it appears that we would be led to life-history theory and the pathological complexity thesis, since an evolutionary approach to life and mind is not failing to consider the autonomy of the organism - instead it shows that autonomy ought not be conceived as the fundamental property of either life or mind. If there is anything like such a fundamental property, it would be the teleonomic trade-offs any Darwinian creature has to deal with in the maximization of their fitness, and it is here, I argue, that we find the origins of consciousness.

How to think about Selfhood?

Despite flaws in the autopoietic approach, the basic idea of consciousness being related to life, self-production, and agency is a useful one and has been influential both in Godfrey-Smith's attempts to naturalize subjectivity and Ginsburg and Jablonka's (2019) recent treatise on the evolution of consciousness. They maintain that consciousness is something like a new mode of being or way of life that has arisen somewhere during the aforementioned Cambrian explosion; a particular animal way of being in which subjective experience is simply part of what it means to live a flexible life. Note, that de Waal did not assign a zero level of self-awareness to any animals in his depiction of a gradualist view of self-awareness (**B** in Figure 2.2). This is because he thinks that all animals require at least a minimal form of a self-concept. But treated literally, this claim - despite being intuitive - is surely

too strong when we think of the less mobile and more 'plant-like' side of the animal branch of life such as corals, anemones, and sponges.

To think that consciousness evolves in organisms that become more recognizably agent-like does not have to mean that a sense of self lies at the origin of consciousness. Self-related capacities might be close to necessary for the lifestyles of animals in the older but still common folk understanding of animals as mobile organisms with a nervous system: a particular animal *mode of being*. Here, a registration of a difference between self and other seems to be a requirement for efficient movement and action-selection in a complex body (see also Godfrey-Smith 2020b). After all, such animals must have some way of distinguishing themselves from the world. As Birch et al. put this point; "[a]ny complex, actively mobile animal needs a way of disentangling changes to its sensory input that are due to its own movements from changes due to events in the world" (p. 797).[8]

But that doesn't necessarily mean that the self itself is experienced. The hard problem challenge appears to hardly be answered here, given that many of these cognitive processes appear to operate unconsciously in humans. Worse, we do not seem to be able to provide an answer as to how sensory and evaluative experience evolve out of a feeling of selfhood, making this dimension look more like a later invention. The problems of the autopoietic tradition appear to generalize to any internalist view of consciousness that doesn't focus directly on feedback between organism and environment, and the role of selfhood for the organism. This doesn't mean that properties related to selfhood, agency, subjecthood, and the like are unimportant, but that it isn't in the experience of a self *per se* that we find the fundamental property of consciousness.

Consider the evolutionary picture offered by Feinberg and Mallatt (2016) in their *The Ancient Origins of Consciousness*, where a distinction is drawn between exteroceptive, interoceptive, and affective sides of experiences. In thinking about the origins of self-consciousness, it does not appear hard to imagine that it simply constitutes a combination of these capacities. The most minimal form of selfhood would constitute a distinction between exteroception and interoception and this would make the evolution of self-consciousness look like some kind of 'disentangling' on the sensory side of consciousness. It should not be expected from the earliest organisms possessing some wash of sensory sensation and hence should be seen as a later layer added on top of this. What we thus need in order to uncover the ancient origins of this subjectivity is a more gradualist picture of that basic idea - in which complex cognitive processes are recognized that, as Godfrey-Smith (2020b) argues, largely go on behind the scenes, unlike the features that we now usually associate with the core of consciousness. The dimension of selfhood is more plausibly seen as something gradually build out of more basic sensory and evaluative experiences, eventually giving rise to a meaningful recognition of a subject and thus *subjective* experience, but it is in the actual building blocks of experience that we must seek the origins of consciousness.

Let us not add to the number of ingredients evolution has to produce in one fell swoop for consciousness to arise, as if (in the example too-often used by creationists) the human eye magically appeared into existence. As Ginsburg and Jablonka (2019) point out, Darwin had trouble conceiving of how the eye could gradually appear, "but as he explained, this was the problem of a failure of his imagination rather than

[8]See also Hurley (1998); Merker (2005); Godfrey-Smith (2016d, 2020b); Trestman (2017).

a failure of his evolutionary theory" (p. 32). A gradualist picture may be wrong, but those studying consciousness have so far spent little effort on trying to make the best possible case for it.[9] A Darwinian picture, while perhaps hard to think about, may be the only path to finding the true place of mind in nature. We can happily embrace the possibility that there could be something like "hemi-semi-demi-pseudo-proto-quasi-minds" (Dennett 1995b, p. 108) that may appear in creatures very different from ourselves. From this bottom-up perspective, the construction of a self, of a subjective point of view, can again be seen as an outer layer in the onion that is consciousness. Shedding it away gets us one step closer to finding the origins of subjective experience.

This finally leaves us with two remaining options and a substantial narrowing of the explanatory gap. While diachronic experience, synchronic experience, and the experience of a self can be understood as features *of* the way conscious experience can be structured, as opposed to constituting it, it remains an open question as to whether the origins of consciousness can be found on the sensory or evaluative side. In the following, I will argue that there are reasons to doubt that the answer is to be found in the sensory dimension, which suffers from what are almost the inverse problems from the dimension of selfhood.

3.3.2 Sensory Experience

A sensory explication of phenomenal experience is widespread among both philosophers of mind and neuroscientists investigating consciousness. The sensory side, and in particular studies on human vision, have often served as the model for *all* of consciousness, likely due to its ties to both of the equally influential notions of a *point of view* and an *awareness of*. It has been seen as the key to understanding phenomenological experience, which was only aided by the fact that even a science of human consciousness has long been deemed impossible and continues to be seen by some with suspicion. To a large extent, the focus on vision was an attempt to make research on consciousness as scientific as possible, since this capacity seemed more readily testable.

Affective neuroscience, which focuses on the moods, personality, emotions, motivations, and feelings of both humans and animals, has continued to be viewed with suspicion; not only among cognitive neuroscientists who saw this research as too closely related to 'consciousness science' (see Cacioppo and Gardner 1999), but also among philosophers who saw this empirical research as largely irrelevant to their conceptual analysis of emotions (see Griffiths 1997, 2017).

The problem with this approach, of course, is the possibility that our early models and theories, developed based on human vision, may influence the way we treat all other aspects of subjective experience. Perhaps our current troubles with naturalizing consciousness are due to path-dependence and we would have been in a much better place if we had begun with the study of affect and valence (see also Solms 2021). But to advance such an alternative approach, I will first argue against the centrality of a sensory model of all experience.

[9] See Lee forthcoming for a recent critique of common arguments against a gradualist view of consciousness.

Experimental Philosophy of Consciousness

One source of evidence against a sensory-centric view has so far been given little attention in the philosophy of mind, and that is experimental philosophy (of mind). In particular, recent work by experimental philosophers such as Justin Sytsma and Edouard Machery on the intuitions of the public regarding their conscious experiences appears to offer a strikingly different picture from the major views within philosophy of mind and one that fits strongly with the evaluation-first view defended here (Sytsma and Machery 2009, 2010; Sytsma 2010; Machery and Sytsma 2011; Sytsma and Machery 2012; Sytsma 2012; Sytsma and Ozdemir 2019).

In an influential study, Sytsma and Machery (2010) show that members of the public do not share the philosophical consensus view that consciousness is characterized by its 'phenomenality'. While there is much debate on how this notion ought to be conceived, there is broad consensus in the literature that "subjectively experienced mental states have phenomenal properties: There is something it is like to see red, smell banana, feel anger, and be in pain" (Sytsma and Machery 2010, p. 324). Ordinary folks, however, do not appear to recognize a commonality between all of these states. When asked whether a relatively simple robot can feel pain, ordinary people are highly skeptical. However, when asked whether the robot can see red, the folk were much more likely to say yes compared to philosophers in the same study. This suggests that the folk does not share the philosophical conception of consciousness and subjective experience, i.e. they do not define consciousness as philosophers do, in relation to their 'felt' or 'phenomenal' properties (Machery and Sytsma 2011). But if they do not rely on the philosophers' 'consensus' view, this raises the question of how they do think about consciousness.

The findings of Sytsma and Machery suggest that the ascription of conscious mental states to a robot rests on a hedonic evaluation, i.e. whether something feels good or bad. Ordinary people appeared to distinguish those states with valence or affect such as pain, moods, and emotions and those without such as a pure sensation of red or the smell of isoamyl acetate (Sytsma and Machery 2010). Ordinary people seem to perceive consciousness as an evaluative experience, thus refusing to attribute mental states to robots when they include experiences that we would typically associate with hedonic evaluation. This would be a radical inversion of the way philosophers have thought about the problem, with evaluation here taking a much more centre stage.

Mere sensory discrimination did not appear to be particularly controversial when attributed to a robot, which is strikingly different from Dennett's (1996) assertion that the robot Cog[3] "cannot yet see or hear or feel at all" (p. 16) [cited in Sytsma and Machery (2010, p. 302)]. And yet, ordinary people seemed resistant to assign a robot the ability to perceive "familiar smells associated with either positive or negative valence" (Sytsma and Machery 2010, p. 318). The common folk-concept of subjective experience thus appears to group "different types of perceptual experiences, bodily sensations, and felt emotions depending on their valence" such that only those sensory processes linked with valence are considered restricted to conscious beings (p. 318). These results are interesting because they may undermine the very foundation for the philosophical resistance of those defending the idea that there is something like an unbridgeable "explanatory gap" (Levine 1983) or a "hard problem of consciousness" (Chalmers 1995):

> The hard problem is typically justified on the grounds that we are acquainted with the phenomenal properties of states such as pain and seeing red and that functional accounts of mental states fail to explain how they can have such phenomenal properties. Our findings challenge the first premise of this argument. Because people do not seem to conceptualize their subjective mental life as phenomenal, it is at least unclear that we are pretheoretically acquainted with the phenomenal properties of our conscious mental states.
>
> – Sytsma and Machery (2010, p. 324)

Philosophers have long religiously maintained that subjective experience is essentially conceived in the same way by both themselves and the public - that they are merely using the folk-concept of consciousness. Sytsma and Machery (2010) list multiple high profile examples of this widespread belief even among naturalist philosophers (p. 320), that I shall repeat here: Alvin Goldman (1993) comes out in support of "the basic integrity of the folk-psychological conception of consciousness and its importance in cognitive theorizing" (p. 364). Ned Block (2004), in his "Qualia" entry in the *The Oxford Companion to the Mind* similarly defends the phenomenological view of consciousness among philosophers as *the* folk view. Patricia Churchland (1988) also accepts that the public hold such a view, but urges us to consider the "outright replacement of the old folk notion of consciousness with new and better largescale concepts" (p. 302). Even Dennett (2005), who likewise wants to develop a scientific account of consciousness by more or less banishing the notion of 'qualia', holds that it is part of our folk conception of consciousness (p. 27). Moreover, the study of Sytsma and Machery (2010) also asked philosophers to make a prediction of the assessments ordinary people would make, which they largely expected to be analogous to their own answers (though slightly less skeptical). But as Dennett's assessment of Cog^3 indicates, philosophers may have been strikingly mistaken by confusing the views of their discipline with that of the public.

If this way of thinking about consciousness, however, rests in a mere artefact of philosophical training, we may have to radically revise our assessment of the common philosophical critique by the likes of Chalmers and Nagel, who have maintained that neuroscientists and psychologists who claimed to have explained consciousness have naively failed to address the 'obvious' hard problem of consciousness.[10] If these experimental results are indicative of the way non-philosophers think about the problem of consciousness, such assertions ought to be seen as quite an uncharitable interpretation of what the scientists are doing and thinking.

Sytsma and Machery (2010) posit that "it might be that like the folk, they do not conceive of subjective experience as being phenomenal, in spite of having plausibly carefully considered [']what it is like['] for them to see red, feel pain, and so on" (p. 323). Both the public and scientists may simply not consider a further phenomenological property or 'qualia' that needs to be addressed, which they think is in line with their own experiences, where "many ordinary people either don't understand or don't take seriously the philosophical concept of phenomenal consciousness even after a lengthy explanation" (p. 323). If the hard problem is only something to be recognized once one has come to be 'indoctrinated' by the orthodoxy in the philosophy of mind, this may explain why some scientists, after passionately endorsing the possibility of and need for a science of consciousness, have later taken on

[10] See Chalmers (1995) and Nagel (1974).

the view that the hard problem cannot or may not yet be solved, and that they are instead focusing on the 'easy' problems of consciousness (see also Crick and Koch 1990; Klein and Barron 2020). This may in some cases be merely a strategic choice to avoid the charges of naivety hurled against those scientists claiming to provide explanatory sketches and theories of consciousness. But if so, it would be an unfortunate one since it in turn legitimizes talk of the hard problem as something that cannot be overcome by ordinary science.

Once a naturalistically inclined scientist interested in developing a science of the mind gets too close to what Dennett (2017a) described as the "Cartesian gravity" of Descartes' dualistic way of thinking it becomes all but impossible to escape once one has orbited too close to "Planet Descartes" (p. 20). But this is what the very mention of the hard problem unfortunately enables. If the vision-centric bias in our thinking about and research of consciousness makes it all-but-impossible to shake off the dualist leftover of what Dennett (1991) described as a 'Cartesian theater' in which everything is consciously presented to a homunculus, we may wish to return to an earlier evolutionary view suggested by Dawkins (1998) in which "[t]he key to the origin of consciousness itself may lie in the emotional experience of suffering" (p. 324). This at least is the core of the pathological complexity thesis defended here: evaluative experience as the original and most basic kind of subjective experience.

Ultimately, Sytsma and Machery (2010) have provided us with a beautiful case for the usefulness and perhaps even necessity of experimental philosophy to progress in philosophical debates muddled by appeals to intuition.[11] Godfrey-Smith (2020b), while accepting that "Sytsma and Machery may be right about the everyday conception of experience", nevertheless maintains that "everyday thinking may also be mistaken" (p. 311). That is certainly correct. However, in a debate that has extensively drawn upon near-certain assertions about the folk concept of phenomenality as a support on which to build our models of consciousness located on the sensory side, this work at least undermines a central argument for betting on this dimension.

Due to considerations of space, I will not offer a detailed defense of why experimental philosophy is useful to philosophers. Firstly, I have done so elsewhere (Veit 2021c; Browning and Veit 2022b), and secondly, so have many others (see Knobe and Nichols 2017 for an overview). Within the context of this thesis it is sufficient to acknowledge that the intuitions of the public may not settle this debate, but that we should at least take seriously the insight of the folk into subjective experience as a source of evidence in judging which dimension of consciousness should be seen as the most fundamental. Let us now turn to my core argument against the sensory side, i.e. an excessive externalist representationalism.

The Problems with Externalist Representationalism

Yet another elegant argument against the centrality of a sensory model has come from Godfrey-Smith (2020b), who despite his insistence on a possible separation between the two dimensions of sensory discrimination and evaluation, has argued that a "problem with much recent work in philosophy is the idea that sensing is not only an important part of experience, but just about all that goes on there"

[11] Though this is not to say that their case for two different concepts of consciousness in the public and among philosophers has gone unchallenged. See Huebner (2012); Talbot (2012); Peressini (2014); Chalmers (2020) for challenges to their results, and a recent empirically supported defense of their view (Sytsma and Ozdemir 2019).

(2020b, p. 113). Indeed, much of the recent work in the philosophy of mind uses the words 'sensing' or 'perception' as umbrella terms to cover all forms of subjective experience. Moods, emotions, and feelings are seen through this lens as the detection of some internal phenomena such as thirst or hunger. The only difference of these from olfaction and vision is a matter of perceptual direction: inward versus outward.

Godfrey-Smith (2020b) uses two influential examples from different generations: Fred Dretske, who influenced Godfrey-Smith's own representationalist views when he was a student, and Jesse Prinz, who was his colleague at CUNY and in turn appears to have made him skeptical of the promises of representationalism. While Dretske (1993) doesn't assert that all of subjective experience must be perceptual, he cites Velmans (1991), Humphrey (1992), and the originator of the influential *global workspace theory* Baars (1988) to support the view that perceptual experience and belief are taken to be the "clearest and most compelling" paradigm cases of consciousness in empirical research (p. 272). In a revealing passage, Dretske argues:

> Why can't we, following Damasio (1994), conceive of emotions, feelings, and moods as perception of chemical, hormonal, visceral, and musculoskeletal states of the body?
>
> This way of thinking about pains, itches, tickles, and other bodily sensations puts them in exactly the same category as the experiences we have when we are made perceptually aware of our environment. The only difference is that bodily sensations are the experiences we have of *objects* in the body (the stomach, the head, the joints, etc.), not *objects* outside the body.
>
> – Dretske (1999, p. 117) [italics added for emphasis]

But it may well be a mistake to think of these experiences as representing *objects* in the body, which makes these accounts feel very much like attempts to somehow objectively experience states of the world as described by Newtonian mechanics. The very reason these accounts haven't satisfied proponents of the hard problem has been the lack of recognition of a *subject*. In order to address these problems, *subjectivity* needs to have a role, and not be ignored as a mere by-product of the existence of sensory representation. But this is next to impossible due to a reliance on a strongly externalist mode of explanation. Rather than recognizing the epistemological straightjacket of a strongly externalist approach, representationalists tend to simply bite this bullet and declare the sense of a self or subject *unimportant*.

At the end of his monograph *The Conscious Brain: How Attention Engenders Experience*, in which he defends his Attended Intermediate-level Representation theory of consciousness, Prinz (2012) confidently asserts that "[a]ll consciousness is perceptual; there is no distinctive cognitive phenomenology or any phenomenal self" despite his own concession that "[almost] all of the empirical research reviewed here comes from vision science" (p. 341). Indeed, his own theory is simply an extension, if not generalization, of his earlier representationalist neurofunctional theory of visual consciousness (see Prinz 2000), with little to no attention paid to conflicting empirical work on other dimensions of consciousness.

Such unashamed confessions of confidence in the centrality of vision as the paradigm of all of consciousness should at least raise some worries that this path may have been something like a wrong turn. Godfrey-Smith sees this as a general tendency of much current information-theoretic work on consciousness - such as Michael Tye's (1995) PANIC theory of consciousness and Stanislas Dehaene's (2014)

version of the Global Workspace Theory due to Baars - treating consciousness as a special qualitative way of information being represented in a mind (Godfrey-Smith 2020b, p. 115). These representationalist theories share too much with an older empiricist view of the mind "as merely reactive, needing to derive its patterning from elsewhere" (Godfrey-Smith 2020b, p. 188). That such a view leaves something out was rightly recognized by internalists such as Thompson (2007): "Representationalism neglects the subjective character of experience" (2007, p. 283) and if "there is to be progress in understanding mental imagery as a form of human experience, and not merely as a form of mental representation, then we need to do better" (2007, p. 267). But to do better, we do not need to resort to internalist views of the mind.

An Alternative to Internalism and Externalism

Godfrey-Smith (2020b) urges us to introspect on whether sensory experience and belief are really the most compelling cases of consciousness. He suggests that what I have discussed under the label of affects in Chapter 2, i.e. "emotions, willings, moods, and urges", seem to be at least as, and in his own case "quite a bit *more* clear, as cases of conscious experience, than beliefs" (p. 114). But even if we were to allow that vision is the most paradigmatic case of human conscious experience, this may simply be an artefact of our evolutionary path, a path that has made us masters at learning about and improving our environments, thus making vision appear to represent the external world 'objectively', like a mirror-image presented in something like a Cartesian theater.

Godfrey-Smith suggests that an "alternative to [the strongly representationalist] view, rather obvious but neglected, is the idea discussed in our section on the experience of selves that a mood is not a *presentation of* some fact or condition; it is just *the way things are* with you, at that time" (2020b, p. 114). That was one feature of the selfhood-first view that made it attractive. However, both the autopoietic tradition and the representationalist tradition fail to recognize feedback between internal and external factors. Treatments of consciousness by Prinz, Dretske, and others neglect the organism's dynamic role in subjective experience, simply treating qualia as little more than representations of an environmental state. Internal goings-on such as hunger are seen as just another kind of environment. No place is given to a subject, there is no center of agency, no feedback between the environment, action, and consciousness, making it unsurprising that such views turn the place of *subjective* experience in nature into something like a mystery. Subjectivity simply slips through the cracks in such a widening of the explanatory gap.

Here, we can follow Godfrey-Smith's suggestion to "reject the un-ecological side" of the autopoietic anti-Darwinian emphasis on autonomy and instead embrace the importance of both input and output, or perhaps more importantly the causal traffic that goes on between both sides and expresses itself in action (2016b, 788). Unlike Godfrey-Smith, however, I argue that this traffic that is important to understanding the function of consciousness is readily explicable from the evaluative side of experience, which is directly tied to action.

Whereas the dimension of sensory experience emphasises an externalist view of consciousness, the dimensions of self-experience emphasise internalism, two kinds of overreaching that make it extremely hard, if not necessarily unsatisfactory, to explicate either side in terms of the other, which is reminiscent of the dilemma the ethologists faced between the vitalist's internalism and the behaviourist's exter-

nalism. Here, evaluation offers us a way out of this dilemma by being inherently dynamic and agential. What makes an external state good or bad depends on the state of the organism, and likewise whether a state of the organism is a good one depends crucially on the environment. Ecological feedback is built into this Darwinian picture of an evaluation-first view from the very beginning, with action and agency being emphasized.

3.4 The Last Dimension Standing: Evaluative Experience

My goal in this chapter was not to provide definite proofs that theories of consciousness centered on the other dimensions must be wrong, but rather to weaken the confidence one may have for thinking of these dimensions as fundamental to consciousness. In peeling off the dimensions that I consider to be later transformations of consciousness rather than its evolutionary origin, it should hardly be surprising that the picture becomes less certain the closer we move to the core of this mysterious onion. The sensory side, in particular, is a dimension that any account that bases the evolutionary origins of consciousness in evaluation must explain as it would otherwise do no better than its contenders. In offering my account of how sentience evolved in Chapter 4, I will also detail a picture of how the representational sensory side of consciousness, as well as selfhood, could have quickly evolved in the Cambrian once a hedonic mode of being was 'set in place'. Whereas theories of consciousness centered on these dimensions appear to explain the other dimensions somewhat inadequately, I will argue that a theory of consciousness centered on sentience will allow us to keep what is best about both approaches.

After all, many of the experimental paradigms discussed in the previous chapter to distinguish conscious from unconscious perception rely on particular forms of learning such as trace conditioning which suggests that evaluation plays a core role in consciousness. Nevertheless, before I turn to valence as a resource for an alternative theory of consciousness so we should also critically assess the evaluative dimension as I did with the other dimensions. Two important challenges will be met: to motivate that hedonic valence i) constitutes a common core to the evaluative dimension, and ii) that it can help us to make sense of the origins, function, and phenomenological complexity of consciousness.

Hedonic Valence as the Core of Evaluation

What is it that unifies the evaluative states we call moods, feelings, and emotions? As Browning (2020b) readily acknowledges in her attempt to naturalize a subjective notion of animal welfare: what we usually call affective states are "extremely heterogeneous states" (p. 164). However, it appears that not only ordinary people see valence as the common denominator of all these 'affective states', but also a growing number of scientists that study them. Indeed, as I mentioned in Chapter 2, many consider valence to constitute a 'common currency' that makes all of these heterogeneous states subjective comparable. While Browning (2020b) defends a Benthamite view of subjective wellbeing as the common currency of affective states, she considers the possibility that our evaluative experience could be more of a con-

struct, rather than an integrated state: "something like health, made up of multiple different components that, though individually real, do not together form any naturally existing state" (p. 164). I like this comparison because it leads us back to the close connection between health and consciousness that motivated the thesis of this thesis. As I have argued in Chapter 1, health is not just a construct - it is an optimal design-response to the species-specific pathological complexity faced by different organisms in their life-histories, with fitness providing a common currency of evaluation. Similarly, hedonic valence can be seen as the common currency of evaluation for an agent's choices between alternative actions.

But despite the agreement among many scientists that valence constitutes a common currency in humans and many other animals *now*, this doesn't necessarily mean that such a currency was present at the very origins of subjective experience, nor that such a 'currency' must exist. While I will defend both claims in the next chapter, I admit that the notion of a 'common currency' in thinking about the very origins of valence can be something of a mixed bag, because it seemingly presupposes the existence of multiple distinct subjective experiences that are to be compared. Talk of a currency may incorrectly suggest that valence evolved in order to make the different aforementioned states comparable - similarly to how real monetary currencies arose only after there were already goods that could be traded in a more efficient way through the implementation of a common currency. Valence could then be seen as having evolved in response to a certain kind of phenomenological complexity on the sensory side and this appears to be a popular view among those endorsing strongly externalist representationalist views of consciousness. However, we have seen that valence cannot readily be explicated within a sensory model of consciousness. The origins of valence are better conceived as the origins of fuzzy action-imperatives, that arose out of something like the vague discomfort Romanes describes, but which then evolved to have richer discriminatory capacities between different states. Or to put it differently, hedonic valence becomes a common currency through the evolution of richer discrimination and thus the origin of sensory representations. This would explain the evolution of sensory consciousness, the distinction between conscious and unconscious perception within an evolved valence-system of the brain, and it seems to fit the folk understanding of consciousness. This nuance aside, however, I consider the term on balance a useful one to understand the origins of consciousness.

An objection to this view would be to insist that even if we think that hedonic valence constitutes a common currency in our decision-making, our own human experience may strikingly differ from other animals, who may lack such a common currency. To this, of course, we can simply respond that pleasure - unlike language or higher-order symbolic thought - does not seem like something restricted to us. Indeed, it may even be more important for animals to possess a common currency of pleasure and pain, since they cannot engage in the same symbolic cognitive processing as us. In humans with a rare pathology making them unable to feel pain, i.e. congenital analgesia (sometimes referred to as congenital insensitivity to pain) early death is common due to a neglect of or inability to detect injuries and diseases (Thrush 1973; Nagasako et al. 2003; Cox et al. 2006). Importantly, the absence of pain should not be confused with the absence of any negative valence. It is a mistake to think of hedonism as referring to just two states of pleasure and pain; these do not exhaust the scope of all valenced states (whether positive or negative).

The experience of pain nevertheless plays a crucial role in developing a concept of self and one's body in relation to its environment. Those with congenital analgesia must exert cognitive effort to think about or actively represent the potential dangers to their body, since they lack a system of 'punishment' that would teach them from their childhood onwards. While this is certainly not easy, and less efficient than actual pain-experience, it can be done. Animals, however, are unlikely to even make it that far without the fast decision-making and learning of important associations that is enabled through experience of negative valence that can be traded off against other needs such as hunger. They would not be able to think about the likely 'harms' of particular behaviours without painful experiences to make these connections. Such evidence is compelling and the next chapter will offer an account of how this capacity may have evolved, but our own experience of hedonic decision-making appears to be at least plausibly be on the right track in regards to a potentially much more ancient felt experience.

Browning (2020b) puts the centrality of valence in our subjective experience vividly: "I consider myself right now and the combination of states I am experiencing – mild hunger, physical comfort in my office chair, slight head pain from a lingering cold, anticipation of my upcoming lunch, some intellectual discomfort from trying to write this chapter, among other states" (p. 169). While these appear to be strikingly different kinds of experiences, we all share the experience of being able to decide affectively based on how these experiences make us feel and trade off against each other. They become integrated on something like a single scale. What different experiences share is a felt evaluative sensation that becomes integrated into a single state and enables us to compare competing interests and motivations. This comparison does not take place in the cognitive or representationalist sense of a calculation, but rather an instantaneous general experience of one's state - a total state of momentary *feeling* in just that sense of the word. Indeed, it is this immediate insight into our daily experience that has driven the common folk understanding of hedonic decision-making. Nevertheless, since this thesis tries to stay clear from putting too much emphasis on both intuition and our own human experience, I want to avoid committing this mistake here myself. My framework should decidedly not be understood as being merely based on *different* intuitions. Instead, I will focus in the next chapter on the reasons for why we should expect the existence of something like a single scale - a real psychological utility function in which such hedonic values are compared - and how it could have evolved. Furthermore, I will provide counterarguments against those who think that there is no such a kind of 'common currency'.

In his work on consciousness, Damasio (1999) similarly centers feelings and evaluation to make sense of consciousness as promoting advantageous actions, but he requires some minimal sensory representation and feeling of a self at the evolutionary origins of consciousness. I do not see these as necessary for minimal sentience since it gets us uncomfortably close to the idea that consciousness 'pops' into existence once there is a certain degree of complexity in self-recognition, sensory processing, and evaluative cognition. Again, such thinking appears to make the qualitative aspect of consciousness something like a mysterious extra ingredient, although his emphasis on embodied cognition is important to make sense of the role of consciousness in nature.

As I shall argue in the next chapter, what I call *Benthamite creatures* have a

nexus of evaluation at their final behavioural common path that enables them to engage in efficient decision-making under the conflicts of different needs, even in the absence of a sense of self or felt representations.[12] Valence plausibly came into existence with a basic feeling of good and bad, without any *felt* sensory richness. As long as the first sentient beings continued to engage in a behaviour that caused them something like a 'plus sign' and changed it up once it became a 'minus' (in the sense of a vague impervious impulse for action, possibly realized through dopamine and other valence-related molecules with a deep evolutionary origin) there would have been no need for the presence of the other dimensions for this to be adaptive. Under very deflationary senses of the term 'representation', such affective decision-making may well be considered as representing two different states: 'good' and 'bad', in the sense of fitness-enhancing (e.g. Carruthers 2018, forthcoming), but such a usage of the term is unhelpful in understanding i) why some representations are felt, and ii) how consciousness becomes more representational. Importantly, the absence of conscious representations does not imply the absence of representations altogether. Most of what goes on in the brain and nervous system is unconscious and many of these processes may well be representational. But before I explicate this account of mine in the next chapter, let us return to the question of how far we have come in narrowing the explanatory gap.

Valence, Qualia, and the Explanatory Gap

Sytsma and Machery see their results as undermining the very idea that there is *anything* like a hard problem or explanatory gap. I think that deflates the problem too much, though I can see the temptation to adopt this position. Unlike Sytsma and Machery, I see the move towards understanding the biological basis of valence and affects as a *naturalization* of the vexing notion of 'qualia' through an alternative non-vision-centric model of consciousness. While Sytsma and Machery raise the possibility that the valence of consciousness is a problem akin to the phenomenological version of the hard problem, they resist the notion that we couldn't explain it in virtue of functional and mechanistic explanations. They consider it to be straightforward to understand that the "hedonic value of a stimulus or a bodily state seems to be an evaluation of its expected value to the organism" (p. 322). There doesn't appear to then be an additional problem of *why* there is valence. Like Solms (2021), I argue that this makes the evaluative side of experience a compelling target for an attempt to bridge the gap between matter and mind.

The pathological complexity thesis is simply an attempt to naturalize consciousness in all of its phenomenological complexity in a squarely Darwinian framework. To have phenomenological experience *requires* hedonic evaluation. To naturalize the puzzling notion of 'qualia' is simply to explain how and why organisms have such a form of evaluation. Phenomenal states that do not appear to fit the evaluative dimension of consciousness can nevertheless be explained within the context of the evolution of Benthamite creatures. With evaluative capacities gaining in discriminatory sophistication that allows for the distinguishing and comparative evaluation of states, we can offer an explanation for why some sensory processes are conscious, or rather *felt*, whereas others are not.

There is thus no need to eliminate the notion of phenomenological experience,

[12] I will explicate this idea in detail in chapter 4.

which Sytsma and Machery think is as flawed as the vitalist's concept of life (p. 322). Admittedly, there is often only a fine distinction between those who try to naturalize a concept and those who seek to eliminate and replace it. Perhaps the only difference is a degree of sympathy shared with the foregoing work of both philosophers and scientists who studied the mind and in particular our phenomenological experience. In thinking about the origins of mind in hedonic evaluation, however, the explanatory gap will no longer appear as large - making qualia less mysterious than they might at first appear. Inevitably, this will involve some reshaping of how our ordinary concepts of both 'mind' and 'matter' conceive of what goes on in organisms. But that is simply what it means to find the place of consciousness in nature. Some aspects of our folk concepts of these notions might be naturalized, whereas others are eliminated, ultimately providing us with a new scientifically informed view of the mind.

3.5 The Spoils of War

The goal of undertaking this war between the five dimensions was to discover and crown the core of consciousness. While we may lack direct paleobiological data that could indicate which dimension has the most direct line of descent to the dawn of consciousness, this chapter has attempted to reverse-engineer the origins of consciousness. Let me summarize which dimensions lost, and why.

I have argued that diachronic experience can be most readily dismissed as a necessary component of subjective experience since it is largely agreed upon to be absent in some animals, all the while granting them subjective experiences (of some kind). What Birch et al. (2020) called the temporality of experience, can be seen as a higher-order feature of consciousness, not something that is likely to be present at its very origin.

Synchronic experience, while more often seen as a necessary component, must not be present at the very origins of consciousness either, since other animals without a strong connection between both hemispheres are likely to have a more disunified experience. There may be situations where disunity might in fact be more functional than highly centralized processing. Both dimensions of unity appear to be features of the way experience can be organized, rather than what makes them qualitative to begin with.

The experience of a self, I have argued, is something that is hard to disassociate from our thinking about consciousness, due to its centrality in our own conscious experience. But in a strongly gradualist picture, it is unlikely to have been present at the very origins, instead *built by* more basic kinds of subjective experience such as a distinction between exteroception and interoception within the dimension of sensory experience.

This left us with two contenders: the sensory and the evaluative sides of consciousness. Sensory experience has long dominated much of our thinking about subjective experience in philosophy and science. Yet, what this dimension seemingly fails to account for was the very *subject* that is so central to consciousness and makes the hard problem challenge seem especially damning to this dimension. It is then a natural scientific move to consider an alternative way of thinking in order to solve the problems of an old paradigm.

By drawing on recent work in experimental philosophy of mind I have undermined the centrality of the sensory side in thinking about the most evident cases of experience. The *feelings side* which includes moods, emotions, and hedonic evaluations, appears to be what drives the thinking of ordinary people about consciousness and thus provides additional support for the pathological complexity thesis I am trying to develop here. There is great untapped potential for building an alternative model of consciousness based on what are sometimes perceived to be the background features of ordinary human conscious experience: moods, pains, evaluations - features that usually only come into the centre when things go great or badly. This alternative way of thinking has unfortunately been largely resisted, yet has striking support from and has been partially developed across economics, animal welfare science, neuroscience, behavioural ecology, animal consciousness science, and ethology. Even if a theory based on the evaluative side of conscious experience will eventually turn out to be false, we are likely to make much greater progress by developing a picture that has received scant attention, yet ought to be considered as *at least* an equal competitor to the sensory-first views that have been modeled on the human phenomenon of visual experience. Unlike rich human-like vision, after all, a basic sense of evaluation - this primordial emotion - may have been present long before animals had any rich capacities to discriminate states of the world.

It is within an evaluative model of consciousness that it becomes straightforward to see how the other dimensions could be built on top of such a capacity. As Panksepp (2005) once argued: "affective experience may reflect a most primitive form of consciousness [...] which may have provided an evolutionary platform for the emergence of more complex layers of consciousness" (p. 32). These outer layers serve to enrich an evaluative mode of being that describes so much of animal life, with further gains in phenomenological complexity being united under a 'common currency'. And this is ultimately what motivates me to take a valence-first view of the evolution of consciousness: unlike the other dimensions, in an evaluative model of consciousness there is no explanatory left-over regarding how the other mysterious properties of consciousness arise. Whereas the strongly externalist and representationalist view of consciousness based on a model of visual capacities fails to account for selfhood and evaluation, the strongly internalist model of of consciousness as self-awareness fails to account for the functional capacities of representation and evaluation. Both fail to put at centre-stage the sensory-evaluative-motor feedback that is so central to a teleonomic Darwinian view of life. Such feedback ought to be at the centre of any theory of consciousness taken as a way of engaging the world as a vulnerable organism with a complex lifestyle that requires evaluation. Whereas defenders of strongly externalist and internalist approaches struggle to respond to advocates of the hard problem, that these capacities could just as well be executed unconsciously by an organism, evaluation fares better with this problem, since it is precisely the felt valence that makes such capacities functional.

Common-sense usage and introspection support the idea that hedonic valence is a fundamental feature of human experience, making it unsurprising that philosophers since antiquity have held similar views about the importance of pleasure and pain. And it is this evaluative dimension that ultimately won the war of the five dimensions. To unearth the evolution of the first evaluative sparks of experience by drawing on the pathological complexity thesis will be the target of the next chapter.

Chapter 4

Pathological Complexity and the Dawn of Subjectivity

Complexity is worth caring about, but not just any complexity.

– Daniel C. Dennett (2017b)

4.1 Introduction

The goal of the pathological complexity thesis is to explicate the evolutionary origins of 'qualia' in evaluation. This idea, while intuitively attractive due to its direct link to fitness and action, is not as straightforward as suggestion by Darwinian thinkers such as Romanes may have made it seem. All life, as I noted in my response to LeDoux in Chapter 2, has organic states and processes that are beneficial or deleterious to it and some means of responding to these. Health is a universal biological measure of how an organism, whether a single-celled bacterium or a bat, succeeds at dealing with the species-specific pathological complexity challenges within their life-histories. But this doesn't mean we should adopt the biopsychist stance and attribute sentience to all life, nor that we should take the stance of LeDoux and deny that such evaluative behaviour has anything to do with consciousness. Both sides fail to recognize a significant transition in evaluative agency.

My argument is not that *any* degree of evaluative agency makes consciousness automatically appear, but rather that an explosion in pathological complexity through higher degrees of freedom comes to be dealt with through an evolutionary transition towards a hedonic mode of evaluative agency, i.e. what I call *Benthamite creatures*. Here, state-based behavioural and life-history theory offers us the ideal agential theory to emphasize the ecological importance of distinguishing biological success from failure, or in other words the adaptive from the pathological, which led to the evolution of consciousness as an adaptation for efficient action-selection. Benthamite creatures evolve in the form of sentient beings with real psychological utility to track this fundamental distinction between health and pathology; which makes the pathological complexity thesis a continuity thesis between properties of life and the mind. How and why this happened will be addressed in this chapter.

Chapter Outline

This chapter is organized into two main sections. In Section 4.2 'Complexity Worth Caring About', I address the objection of why it should be pathological complexity, rather than any other measure of biological complexity, that should matter for the evolution of phenomenological complexity. In Section 4.3 'The Cambrian Explosion in Animal Complexity', I seek to locate the origins of valence in the computational explosion of pathological complexity that occurred during the early Cambrian. Finally, Section 4.4, 'Conclusion and Further Objections' will summarize the conclusions of this chapter, respond to potential objections, and sketch how the next chapter will explicate the pathological complexity thesis in more detail to naturalize the remaining dimensions of subjective experience within a model of valence as the most basic kind of 'qualia'.

4.2 Complexity Worth Caring About

At the beginning of this section, I have placed a quote from a recent talk Dennett (2017b) gave at an Animal Consciousness conference at NYU. In this talk, he criticized the tendency among philosophers and scientists working on consciousness to only take the first step in explaining consciousness by analysing some of its properties, but leaving what he calls the 'hard question' about what consciousness does *for* the organism entirely unaddressed. Agreeing with the widespread view that an understanding of pain and suffering would get us closer to understanding consciousness, Dennett nevertheless maintained that to understand the evolution of suffering, one has to ask *why* suffering matters. It cannot just be *intrinsically* bad, he said, maintaining that we must give an evolutionary explanation of why it matters to the organism. In order to explain the evolution of subjective experience, we must take a deep dive into the ancient origins of valence and examine what made a conscious mode of evaluation worth having. My answer to this question lies in an explosion of complexity that took place during the Cambrian explosion. But as Dennett emphasizes, we must answer which kind of complexity is worth caring about; it cannot be just any kind of complexity. That the answer to this teleonomic question is the normatively loaded sense of *pathological complexity* will be defended in this section.

In trying to reconstruct the possible evolutionary origins and function of consciousness, we are engaged in the paleobiologist's effort of making sense of a trait by connecting its extant 'users' with its historical traces. When we think about valence in humans, it is choice, desires, motivation, and preferences that come to mind; and it is the evolution of such capacities, related to action, that we will have to pay attention to if we want to understand the dawn of subjectivity. One important figure - who has attempted to develop a plausible natural history of the evolution of human agency and cognitive states resembling the standard folk-psychological states of belief and desire - has been Sterelny (2003), but he unfortunately paid comparatively less attention to desire-like states, a lack that has also been criticized by David Spurrett (2015). Both Sterelny and Spurrett draw on the 'ancestor' of the pathological complexity thesis to make sense of the evolution of preferences: Godfrey-Smith's (1996a) aforementioned *environmental complexity thesis*. To briefly restate it: "The function of cognition is to enable the agent to deal with environmental complexity" (Godfrey-Smith 1996a, p. 3). This was an attempt to make tenable within the mod-

ern framework of evolutionary theory earlier ideas from John Dewey and Herbert Spencer about the continuity between life and mind: the mind seen as a natural consequence of the evolution of living complexity.[1] Unlike Spencer and Dewey, however, who intended to include consciousness in their explanation of mental complexity in terms of living complexity, Godfrey-Smith restricted himself to explaining only basic cognitive capacities, excluding subjective experience. While the pathological complexity thesis differs both in its explanandum and explanans, it is indebted to the elegant explanatory framework and naturalist ambition of Godfrey-Smith's thesis. Furthermore, we may well ask whether environmental complexity could explain the evolution of valence as opposed to pathological complexity. After all, environmental complexity is routinely implicated in the evolution of desire-like states, behavioural flexibility, or for that matter, organismal complexity.

However, while the environmental complexity thesis has an explicit link to action, it was (at least originally) designed as an externalist and representationalist theory. In that respect it shared more with Spencer's externalism than it does with Dewey, who saw the complexity of the mind as something that evolved to deal with problems emerging in the dynamics between organism and environment (Godfrey-Smith 1996a), which is closer to the pathological complexity thesis. Explications of the environmental complexity thesis have tended to pay very little attention to the organism as a "design and control architecture", instead treating the mind as something that decides what to do with the body, conditional on a given external state of the world (Spurrett 2020, p. 5). While this thesis is largely concerned with consciousness, rather than cognition, the two are tightly linked so it is perhaps unsurprising that I resist such a strongly externalist picture for cognition as well as consciousness. Godfrey-Smith's earlier motivation for a teleonomic theory of mind was similarly motivated by tying together externalism and adaptationism, which while admittedly common, should not be accepted as a necessary marriage. A teleonomic view that tries to make sense of the mind within a theory of the organism will have to pay much more attention to the features *of* the organism.[2]

In making the emphasis on the organism clearer, it is useful to mention a largely inverse version of Godfrey-Smith's environmental complexity thesis that has been popularized by Keijzer (2015; 2013). Whereas Godfrey-Smith's account of the origins of cognition largely idealizes away the organism, because of its inspiration from the externalist strategies of behavioural ecologists and evolutionary biologists, Keijzer's approach is inspired by the more internalist traditions of neuroscience and developmental biology. Keijzer et al. (2013) proposes that the early function of the nervous system was to enable action "as a *single multicellular unit*" (p. 68). In this picture, sensing - and thus the input-output story - is relatively unimportant when we are looking at the very origins of the nervous system. The nervous system here largely plays the role of coordinating the body, irrespective of what goes on outside, going hand in hand with the evolution of contractile tissue (muscle) close to the skin or epithelium of an animal. Here, the nervous system evolves in order to solve a non-trivial control problem that has to be re-invented at a multicellular level. In a later paper, written as a direct response to Godfrey-Smith's earlier work on the environmental complexity thesis, Keijzer and Arnellos (2017) explicitly describe

[1] See also Godfrey-Smith (1996b,c, 1997).

[2] Recall in Chapter 1 the criticism by Lorenz of the behaviourists' externalism, that it leaves out everything that makes an organism an organism.

their hypothesis as an internal complexity proposal: "acquiring the fundamental sensorimotor features of the animal body may be better explained as a consequence of dealing with internal bodily—rather than environmental complexity" (p. 421). Despite the apparent conflict between these views, however, one need not necessarily see them as competitors. Spurrett (2020), for instance, argues that we could see them as two versions of a more general view, merely differing in their emphasis.

But what would a general version of such a view imply? *A* complexity thesis? This much, we already knew. Complexity, as Dennett noted, matters, but what kind of complexity? Spurrett (2020) proposes the "friendly amendment of the ECT" that takes into consideration both internal and external sources of complexity: "The function of cognition is to enable the agent to coordinate its (possibly complex) capacities, which can include coordinating those capacities with environmental complexity" (p. 5). But this definition only takes us half of the way. It's not enough to describe what cognition does on a general level. Like Dennett, we should ask the hard question: and then what happens? Why coordinate? Why act? It cannot be complexity *per se*.

In more recent reflections on his earlier work, Godfrey-Smith (2017b) admits that he was overly eager to state the environmental complexity thesis in externalist terms and recognizes that it isn't environmental complexity, *per se*, that matters, but rather the complexity faced *by* and mattering *to* an organism. Here, we ought to reject the dilemma between externalism and internalism. In the context of admitting his erroneous bet on externalism, the extensive discussion of Keijzer in Godfrey-Smith (2020b) may be seen as an attempt at self-correction and taking internalist views more seriously. He explicitly acknowledges Keijzer's influence in rejecting the mainstream representationalist thinking in the philosophy of mind, describing his ideas as an "emphasis on the *shaping of action*" (p. 59) that is so important if one wants to understand the branching of an animal *way of life*.

While Keijzer and his collaborators are a little too caught up with internalist ways of thinking, they are correct to locate the origins of mind in the control of action, rather than as Sterelny (2003) argues in the evolution of rich sensory detection-systems. A Darwinian approach to the mind must ultimately emphasize dynamic feedback between organism and environment, since it is here that a kind of complexity dynamically emerges that *matters* and this can only be found in a teleonomic measure of complexity based on a theory of the organism, i.e. the *dynamic* life-history strategy of the organism. The problem we are faced with is simply the pathological complexity described in state-based behavioural and life-history theory. Features such as environmental and bodily complexity must be seen as variables that are relevant for the pathological complexity of an organism - they may be sources for it - but in asking for the complexity that matters to the organism, it is ultimately pathological complexity as *the* teleonomic measure of biological complexity that we must focus on. In asking for what kind of complexity matters for biological agents such that the capacity for suffering is worth having, Dennett offers an elegant answer:

> The complexity of an autonomous, self-protecting, self-advancing (but mortal, vulnerable) bit of machinery gives us an explanation of why it is equipped to suffer, and why its suffering matters *to it*.
>
> – Daniel Dennett (2017b) [italics added for emphasis]

It is only in a system that is (i) capable of 'experiencing' biological damage from pathological states in the sense of 'suffering', and (ii) capable of acting to avoid such states, that it is worthwhile to be capable of *experience suffering* in the subjective senses of these terms. Inaction toward avoiding biological harms is itself pathological and gives rise to a new dimension of pathological complexity that can help us to make sense of why Benthamite creatures evolved. It is this teleonomic complexity that matters for organisms having to make decisions that lead to biological success or failure: how to avoid pathological behaviour and perform the right action at the right time? It is here that I shall argue that consciousness comes to perform its adaptive rationale.

4.2.1 Pathological Complexity and the Need for Valence

The pathological complexity thesis maintains that the function of consciousness is to enable the agent to respond to pathological complexity, which is the trade-off problem faced by all organisms as they pursue their goals. While pathological complexity constitutes a general design problem on the timescale of organismal evolution, the evolution of behaviour makes pathological complexity a constant problem for organisms to solve throughout their own lives, and it is within this explosion of pathological complexity in the evolution of multicellular action that I shall argue hedonic valence finds its origin.

Out of the many thinkers who have thought about the evolution of consciousness, Dawkins perhaps came the closest to articulating the kind of teleonomic complexity thesis that I defend here. Consider the following illustrative quote:

> Animals usually have more than one kind of danger to avoid. They have *complex tradeoffs* at all levels in order to minimize reductions of fitness in facing a wide range of threats. At different times of the day or year, or depending on external circumstances, they will reallocate priorities: For example, animals may depress or enhance their immune responses, increase or decrease their physiological "stress" responses, or find some stimuli more or less aversive.
>
> – Marian Dawkins (1998, p. 322) [italics added for emphasis]

Like our focus here on evaluative feelings, Dawkins (1998) suggested early on that the very "key to the origin of consciousness itself may lie in the emotional experience of suffering" (p. 324). Notice that Dawkins speaks here of an experience *of* suffering - by which she means physical suffering in the sense of ill bodily health - rather than suffering *as* a mental experience, which can be interpreted as the suggestion that consciousness evolves first and foremost to respond to threats to health. Trained as an ethologist under Tinbergen at Oxford, Dawkins has been one of the most fervent critics of the lack of evolutionary thinking within animal welfare science. In an influential paper with the title 'Evolution and Animal Welfare' in *The Quarterly Review of Biology* Dawkins (1998) argued that to truly understand the welfare and subjective experience of animals, we must use an evolutionary approach as we would for any other biological phenomenon, "[a]nimal welfare, in other words, needs a dose of Darwinian medicine (Nesse and Williams 1995)" (p. 305). Indeed, that there could be a strong evolutionary link between physical and mental suffering, or health and negatively valenced experience, has long been a central tenet of those working in the sphere of evolutionary medicine.

Nesse and Williams (1995), for instance, argued early on that some of the emotional states we disvalue and consider to be indicative of poor wellbeing - such as pain and fear - are evolutionary adaptations that are "unpleasant by design" (p. 26). Consciousness itself can be seen as just such an adaptation to ensure the health of the organism. Evaluative agency evolved in order to deal with the economic decision-making trade-offs of a new and flexible animal way of being. What Dawkins (1998) highlights is the need to recognize a teleonomic notion of "complexity of an animal's adaptive response to various dangers" (p. 322), to which one should also add opportunities. Just as life is an evaluative and goal-directed activity, so is consciousness an evaluative and goal-directed way of engaging with the world, evolved within the context of life. It has evolved in order to respond to pathological complexity, which includes both opportunities *and* problems - such as the possibility for a common brushtail possum (*Trichosurus vulpecula*) to steal an unsupervised fledgling from a nest. What presents an opportunity for the possum also presents both a problem and a danger to the chick, and this ecological feedback between different organisms increases pathological complexity once more.

But we need to be careful not to become too tempted by the representationalist modes of thinking that turned the sensory side of subjective experience into the mainstream model for consciousness. Despite his earlier criticism of the environmental complexity thesis for tying adaptationism and externalism together,[3] Sterelny (2003) maintains that the environmental complexity thesis offers us something like a useful coarse-grained abstraction for investigating the origins of desire-like states. This is largely explained by his interest in the evolution of proto-representational states, and at least in this sense Sterelny is firmly connected to older mainstream representationalist thinking in the philosophy of mind. Spurrett (2015) likewise operates in a strongly representationalist model of the mind and considers the evolution of preferences and common currencies as value representations, though he at least recognizes that the problem of coordinating the body around action is very difficult and much neglected. Nevertheless, it is a mistake to tie the origins of valence together with the origins of representational richness, since it underestimates the importance of efficient decision-making and action control in the evolution of any system with higher degrees of freedom. In the absence of efficient action-control, richer sensory capacities are simply a cost that is not worth paying.

Hedonic valence, as a commanding sensation, is plausibly much more simple than the cognitive processes we associate with sophisticated perception, possibly arising as something very primitive and immune to the demands of Lloyd Morgan's canon, but not so simple as to make it a default for all evaluative processes of life. By taking a design stance, i.e. by explicating the pathological complexity of different organisms and thinking about the properties that would make valence worth having, it is easy to see which properties would be relevant. Vulnerability and mortality *matter*. If a system is 'indestructible' and almost immune to dangers posed by its environments there is little sense in demanding pain, or for that matter pleasure. Autonomy and sufficiently flexible behaviour likewise *matter* because valence evolved to deal with the complexity of choice-problems. A system that cannot respond adequately to dangers or injuries, however, does not appear to require the machinery for evaluation. A capacity for negative valence has little to do with representing the world, not even internal states, and much more with enabling efficient adaptive

[3]See Sterelny (1997).

behaviour that matters to the whole agent. The task of the pathological complexity thesis must be to turn this vague, but popular idea into a precise scientific hypothesis and framework for an ethological science of consciousness.

How organisms ought to deal with their species-specific pathological complexity can be explicated in terms of a unified teleonomic state-based and behavioural life-history theory of organisms that accounts for all the actions an organism can take. But Mangel and Clark (1986) rightly note that anything like a unified foraging theory will become almost impossible to assess, since "more complex models can rapidly become computationally unwieldy" (p. 1135). Typically, behavioural ecologists constrain the option-space of different actions organisms can take to a manageable set. But this is simply an idealization to make the life-history trade-offs manageable to model. Once organisms can take alternative actions that change their place in nature, we are faced with a dynamic programming problem, and "[i]t is well known that dynamic programming problems become computationally infeasible as their dimension increases" (Mangel and Clark (1986), p. 1128). We are faced with a combinatorial explosion in trying to model to optimal life-history strategies for such organisms with high degrees of freedom. But the very reason it is so hard to model the maximization problem of their life-history strategy, is why valence evolved as a proximate common currency for action selection. Pathological complexity is an optimization problem for organism and modeller alike: it is the complexity that matters for the *organism* both in the sense of a subject and object of evolution.

Within biology, some types of dramatically fluctuating environments are thought to favor very simple organisms: if it is very hard to survive bad seasons, the best option may be to have a capacity to reproduce very quickly when times are good (Bonner 1988, p. 49). Godfrey-Smith's environmental complexity thesis neglects this: one way to deal with environmental heterogeneity is to become simpler and reduce the *pathological complexity* an organism has to deal with. To idealize away important features of the internal complexity of organisms in the context of understanding the function of mind forces us to neglect some of the most important features of what makes the organism an agent. In order to understand these complex dynamics of a teleonomic system a state-based approach is needed that pays close attention to the life-cycles and goals of organisms.

If we follow the ethologists' demand to study adaptive value alongside of mechanisms and developments, we must answer the black box problem of how organisms optimize their behaviour. How organisms ought to deal with trade-offs between different goals is in principle no different from how we think conscious agents ought to resolve their conflicting goals. As Okasha (2018) notes, agency in folk psychology, economics, and evolutionary biology typically requires "unity-of-purpose" or at least consistency among goals. We can usefully describe a system as an agent if there is a goal that all the processes and mechanisms work towards. The goal of organisms is ultimately reproduction and the pathological complexity thesis can thus describe them as agents with fitness providing a common currency for evaluating the importance of their different needs. However, as Samuelson and Swinkels (2006) rightly argue, organisms cannot just represent their fitness-function to achieve their goal of reproductive success, since the complexity and lack of informational transparency of their situation makes it impossible "to make the agent a *perfect* information processor" (p. 139). Natural selection was constrained to come up with utility functions over actions in a variety of life situations that in many cases will not directly

map onto fitness, but nevertheless, function in very analogous ways. Edmund Rolls (1999) made a useful distinction here between the kinds of choice mechanisms that are more fixed and found in organisms like plants, and those choice mechanisms that are sensitive to learning in the achievement of a goal. Nature is not transparent and the values of different actions have to be learned (and often *un*learned), which is why Dawkins (2001) think that Rolls' distinction can help us to better make sense of the role of consciousness in the choice problems faced by animals.

What both biopsychists and LeDoux (2019) fail to recognize here is a gradual evolutionary transition in agency: agency evolving as a natural phenomenon, rather than just a general property of all organisms with the goal of maximizing fitness.[4] To learn about the workings of internal mechanisms that achieve this end of teleonomic action selection, behavioural ecologists will need to recognize that we need something like a common currency of fitness to compare different actions:

> Any attempt to understand behavior in terms of the evolutionary advantage that it might confer has to find a "common currency" (McFarland and Sibly 1975; McCleery 1977) for comparing the costs and benefits of various alternative courses of action.
>
> – McNamara and Houston (1986, p. 358)

As I previously mentioned, common currency claims are common in the behavioural sciences, and Rolls (1999) similarly maintained that "a common reward-based currency appears to be the fundamental solution that brains use in order to produce appropriate behaviour" (p. v). Furthermore, several scientists have linked this idea to consciousness, including Merker (2007), Ginsburg and Jablonka (2019), and Solms (2021). The suggestion by Panksepp (2005, 2011) to focus on what he called the SEEKING system as an evolutionarily old but conscious motivational pull involved in the foraging activities of animals can also be seen as such a view. Likewise, Humphrey (2011) has argued that the origins of consciousness might lie in the intrinsic valuation or 'enjoyment' of continued survival, an idea that admittedly fits well with our notion of Benthamite creatures.

However, the view I defend here of minimal hedonic valence at the 'dawn of qualia', owes the most to Michel Cabanac, who has unfortunately received comparatively little attention, despite having arguably done the most to link the idea of a common currency to consciousness. Indeed, he has perhaps been the most outspoken contemporary defender of the old utilitarian *Benthamite* idea that this proximate common currency is the hedonic experience of pleasure and pain. This, he argues, is implicated in the early evolution of sentience in the early Amniota, which are estimated to have split from the rest of life around 330 million years ago (Benton and Donoghue 2007). Together with his collaborators, Cabanac has been one of the first to emphasize the importance of positive and negative feelings in decision-making trade-offs in both humans and non-human animals (Cabanac 1971, 1979; Cabanac and Johnson 1983; Cabanac 1996, 1992; Balasko and Cabanac 1998b,a; Cabanac 1999; Cabanac et al. 2009). This is a different kind of common currency claim from that of McNamara and Houston, which is not about the problem of how behavioural ecologists ought to model the economic problems faced by organisms, but rather of

[4]This is also why I have criticized Okasha's 2018 monograph on agency as a concept in evolutionary biology, for having little to say on the actual evolution of agency as a real phenomenon in nature (Veit 2021d).

how organisms themselves deal with their economic trade-offs. Here, the common currency is a real psychological state:

> In natural settings, the goals competing for behavior are complex, multidimensional objects and outcomes. Yet, for orderly choice to be possible, the utility of all competing resources must be represented on a single, common dimension.
>
> – Shizgal and Conover (1996, pp. 37-38)

While Shizgal and Conover do not argue that such a common currency must be conscious, one can readily see why Cabanac connects such common currency claims to argue that hedonic valence will be able to function as a mechanistic proxy that mirrors the structure of the fitness-maximization problem space, and this is why the work of Cabanac is perhaps the most important inspiration for my attempt to find the origins of consciousness in hedonic valence.

Such a proximate common currency adds functional value by making the complexity of the computational problem tangible, enabling organisms with a large behavioural option space to weigh alternative courses of action against each other. Organisms are often faced with what microeconomics studies as a so-called 'substitution problem'.[5] Some needs and motivations are substitutes, i.e. one can be satisfied (at least partially) by satisfying the other. Others, such as sleep and foraging, conflict and need to be evaluated against each other in terms of importance. Benefits of one action need to be computed against the costs of foregoing another and the difficulty of this pathological complexity is rarely given enough appreciation, as if accurate representations of the world alone could fuel adaptive success.

Both Godfrey-Smith and Sterelny - in their emphasis on environmental complexity in the evolution of mind - have neglected this 'internal' source of teleonomic complexity and underestimate the difficulty of achieving efficient action selection (see also Spurrett 2020). This is why the pathological complexity thesis treats phenomenological complexity as something that must ultimately be subservient to evaluation, since it is here that it becomes discharged in action. Like Cabanac, I maintain that pleasure and pain are central in the evolution of animal consciousness, though I will argue that the evolutionary origins of this capacity are quite a bit older than he suggests, since i) - as I have argued in the previous chapter - we should not confuse the very origins of valence with its later function as a common currency, and ii) his argument that the increased complexity arising from terrestrial environments for amniotes gained made this capacity worth having underestimates the pathological complexity of underwater life ever since the Cambrian.

4.3 The Cambrian Explosion in Pathological Complexity

> *[I]t seems certain, as a matter of observable fact, that the association of Pleasure and Pain with organic states and processes which are respectively beneficial and deleterious to the organism, is the most important*

[5] See Shizgal's response to the question of whether there is a common currency for all sensory pleasures in Kringelbach and Berridge 2010, p. 18).

> *function of Consciousness in the scheme of Evolution. And for this reason I have placed the origin of Pleasures and Pains very low down in the scale of conscious life.*
>
> – George John Romanes (1883, p. 111)

If dealing with pathological complexity is the raison d'être of valence, it is tempting to examine the history of life for explosions in complexity that plausibly made this capacity worth having. Doing so changes consciousness from being just a problem biologists may or may not want to address as an explanandum, to an explanation for a rise in biological complexity itself. The most rapid and puzzling explosion of complexity in the history of life is the Cambrian explosion 541 million years ago. In the introduction to this doctoral thesis, I have highlighted the importance of the *Cambrian Explosion* for the evolution of animals, since it is here that we observe the origin of all the major animal phyla we see today (Maloof et al. 2010). Stephen J. Gould (1996) even argued that it constituted the highest degree of diversity in animal life forms, making it maximally *disparate*.[6] The label 'Precambrian' emphasises the importance placed on the Cambrian period; seemingly representing the evolutionary equivalent of the Christian practice of identifying all time prior to the alleged birth of Jesus as 'BC'.

What caused this explosion of complexity, however, remains contested. This lack of a satisfying explanation has led some scientists and philosophers to seriously consider the possibility that we may be able to feed two birds with one scone, by suggesting subjectivity, agency, and other capacities related to consciousness as a (partial) explanation for the Cambrian explosion (see for instance Trestman 2013; Feinberg and Mallatt 2016; Godfrey-Smith 2016a; Ginsburg and Jablonka 2019). If they are right, this would be an early birth of consciousness, indeed. Over the 'short' timespan of the 20 million years that followed, the Cambrian explosion led to complex multicellular body plans, nervous systems, behavioural repertoires, and modes of sensing.

4.3.1 A New Mode of Being

What we see in the Cambrian is the emergence of a new animal lifestyle in which agency and subjectivity come to play a crucial role: a "different mode of being" (Godfrey-Smith 2020b, p. 79).[7] It is these features that come to the mind of most people when they hear the word 'animal': active, sensing, and mobile creatures. Indeed, some think that all such beings are sentient. The folk usage of the term 'animal' as a reference to living multi-cellular entities capable of goal-directed movement is thus not completely misguided when it is used to refer to something like a different *mode of being*. Aristotle, who was not aware of the microbial world of life where these capacities can also be found, used the properties of motility, sensing, and goal-directedness to distinguish a special animal mode of being from those of plants, whose mode of being consists in self-maintenance, growth, and reproduction. He called this animal mode of being the 'sensitive soul', to be distinguished from the merely 'nutritive soul' of plants and the 'rational soul' that humans possess in addition to the other two (Aristotle 1991), which influenced Ginsburg and Jablonka

[6] See Sterelny and Griffiths (1999) for a critical discussion of Gould's views.

[7] See also Ginsburg and Jablonka (2019).

(2019) to title their book *The Evolution of the Sensitive Soul*. This way of thinking about animals treats them as possessing something *extra*, rather than a mere branching in the tree of life. While LeDoux (2019) is right that even single-cell organisms exhibit goal-directed survival behaviour and sensing, he fails to recognize that there is a significant transition in the evolution of a distinctive animal lifestyle, and it is here that I locate the origins of sentience similarly to Ginsburg and Jablonka (2019) or that matter Godfrey-Smith (2020b).

From an evolutionary perspective, the Metazoan branch of life is much older than the Cambrian, plausibly already branching off from the rest of life 800 million years ago in some more or less recognizable transition towards multicellular individuality.[8] During the Ediacaran, which began roughly 635 million years ago and ended with the Cambrian, we find the first definite animal fossils, but they are largely plant-like and their behavioural capacities were simple (Peterson et al. 2008). The ancestors of such lifestyles are, of course, still around us now. Godfrey-Smith (2020b) vividly describes how scuba-divers will inevitably encounter something like a 'breathing forest' when encountering a 'garden' of sponges, corals, and anemones, which are located somewhere between a plant and animal *lifestyle*, yet do belong to the animal branch of life.

In the evolutionary scenarios advocated by the aforementioned defenders of an early view of the dawn of consciousness, much of the focus has been placed on interactions *with others*. This makes a lot of sense in a sensory-focused view, since it is here that we see the evolution of sophisticated eyes and 'tools' such as claws, for the engagement with other organisms. Interaction between *subjects* starts to *matter*; movement can become both 'flight' and 'attack'. In their response to commentary of mine that first introduced the pathological complexity thesis (see Veit 2022b), Merker et al. (2022) interpret pathological complexity as just this emergence of interaction and co-evolutionary arms races in the Cambrian:

> Veit proposes that consciousness arose as a means for organisms to deal with what he calls pathological complexity. We assume that what he has in mind is the kind of complexity that arises in coevolution and evolutionary arms races, say of the predator–prey kind, which became acute with the evolution of large, image-forming eyes, hence his reference to the Cambrian Explosion.
>
> – Merker et al. (2022, p. 55)

While this new dimension of interaction certainly leads to another explosion in pathological complexity by making the life-histories of organisms vastly more complex, it is here that I locate the evolution of sensory experience, rather than the very origins of consciousness. As I shall argue, avoiding the overemphasis on interaction and sensing is a key distinguishing factor between my approach and competing views regarding the evolution of consciousness during the Cambrian. Taking a look at these competitors will help illuminate this difference.

Ginsburg and Jablonka (2019) see the Cambrian explosion as the driver of what they call *unlimited associative learning* (UAL): a special form of associative learning with a vast openness for new complex behaviour, that they consider a transition

[8] While I have previously written about the challenge of evolving multicellularity (Veit 2019), the evolution of multicellular agency constitutes a distinct transition of its own.

marker for the presence of consciousness.[9] This is because they think that UAL ties together eight features of consciousness widely acknowledged among philosophers and neuroscientists: 1) global accessibility and broadcast, 2) binding/unification and differentiation, 3) selective attention and exclusion, 4) intentionality, 5) integration of information over time, 6) an evaluative system, 7) agency and embodiment, and 8) registration of a self/other distinction (Birch, Ginsburg, and Jablonka 2020, pp. 55-56). The details of this list do not matter much for the goals of this thesis and while I readily acknowledge that these features are important for the *shape* of consciousness, like Godfrey-Smith (2021b) I am not convinced of the idea that these features all need to appear together for consciousness to emerge.[10] While Ginsburg and Jablonka (2019) try to find minimal hallmarks of consciousness by excluding features unique to humans such as language, or at least only found in rudimentary forms in other animals such as meta-cognitive representations, we are still faced with the worry that the remaining features could exhibit a human-centric bias. As my discussion in Chapter 2 and 3 hopefully made clear, in order to avoid the idea that we must explain consciousness in terms of a certain rich human form of experience, for which all properties (whatever they are) must be there, or otherwise take the organism to lack subjective experience, we are in need of a comparative bottom-up approach that challenges combination-views. The diversity of life should be reflected in very different ways of experiencing the world, so I am skeptical of putting too much emphasis on a certain combination of features, that may be arranged in very different ways. UAL may well constitute a marker for an evolutionary transition of consciousness towards becoming more recognizably human-like, but the very basis of consciousness is more plausibly found in one of its properties, rather than the combination of a variety of capacities that have transformed consciousness across evolutionary time. For this, I emphasize what they list as their sixth hallmark: an evaluative system. The evaluative ability to avoid harmful stimuli and seek out beneficial ones is supremely important: survival matters. And some basic capacity for an imperative plus or minus 'feel' can readily play important adaptive roles prior to any form of combination of the above-mentioned 'hallmarks'.[11]

Others, such as Feinberg and Mallatt (2016), also offer an account of the origins of consciousness in the Cambrian, though their emphasis is on the evolution of eyes and exteroceptive consciousness as the original source of consciousness, which makes sense if one locates the origins of consciousness in the sensory dimensions. These approaches have so far failed, as Merker et al. (2022) rightly note in response to the pathological complexity thesis, to address the challenge of why "conscious vision, rather than simply better visually based performance operating unconsciously, is needed to meet the transition's functional challenge" (p. 55). My answer is that the origins of consciousness lie in the functional role of dealing with the complex trade-offs arising from the earlier explosion in pathological complexity due to the demands of controlling a multicellular animal body.

Lastly, as I noted before, Godfrey-Smith (2020b) primarily emphasizes agency and subjectivity, but these do not constitute a single property. They constitute

[9] See also Birch et al. (2020).

[10] See also my essay review of their monograph for a more detailed analysis and critique of their account (Browning and Veit 2021a).

[11] See also Barlassina and Hayward (2019); Barlassina (2020), who try to offer an imperative metaphysics of valence that, while I don't agree with all their arguments, nevertheless supports the picture offered here.

a variety of capacities than can be described as making organisms more agent-or-subject-like in some respects. While a detailed evolutionary journey from more object-like organisms to genuine conscious subjects will inevitably involve the gradual evolution of subjective experience and make consciousness less mysterious, it tells us little about the raison d'être and very origins of consciousness, unless we investigate the evolutionary origins of capacities that make organisms more subject-like.

While I think that all of the above approaches are important for understanding the evolution of phenomenological complexity, my problem with all these theoretical proposals for the evolutionary origins of consciousness is that they take a starting point at which animals have already gained a degree of complexity that I argue is indicative of a transition of consciousness towards more complex representational capacities. They all already assume some basic capacity for action and sensing as a given, which they then argue leads to interaction driving an arms-race in which subjective experience makes sense. But in all this focus on *inter*action, it is lost that action itself constitutes a major problem (as the work by Keijzer and colleagues nicely demonstrates). And it is the solution to this problem that caused the Cambrian explosion.

4.3.2 Action!

Some clarificatory remarks on my usage of the term 'action' will be useful here, since it is in the evolution of action that pathological complexity explodes and evaluative experience arises. Following Spurrett (2020) and the ethologists, I treat 'action' here in the teleonomic sense of any kind of functional activity biological agents produce in the usage of their degrees of freedom, rather than (as is common in much of philosophy of action and mind literature) as an exclusive term for intentional behaviour.[12] This view of action is deliberately broad and not meant to provide a clear dividing line between adaptive activities of organisms and what we might typically think of when we hear the term 'action', since their difference is i) only a matter of degree, and ii) my usage of this term is meant to capture boundary cases and include minimal senses of action, such as can be seen in plants producing chemical defenses. It is these boundary cases that we need to pay attention to if we want to understand how organisms can become gradually more agent-like over evolutionary time without thereby implying that there is no difference in degree between the 'actions' of a plant and those of an arthropod. Alternatively, we may wish to describe boundary cases in plants, such as the secretion of defensive toxic molecules and growth towards the sun as 'action-like'. The terminology of 'degrees of freedom' is especially useful here to clarify my argument since it helps to recognize that while plants can engage in a variety of adaptive activities, these are often not at the exclusion of others. Actions in the sense I am interested in here, are following Spurrett (2020) necessarily at the exclusion of others (which is admittedly often a matter of degree). Without being able to engage in multiple activities at once, organisms with high degrees of freedom have to engage in vastly more trade-offs and thus exhibit much higher pathological complexity.

I am thus sticking closer to work in robotics, artificial intelligence, and cybernetics, where the computational complexity of building a teleonomic system is readily

[12]This parallels the use of 'behaviour' as functional activity by Millikan (1995).

recognized. After all, computers are faced with their own pathological complexity in terms of optimal design and the teleonomic obstacles to the achievement of their ends. The usage of biological terms such as 'viruses' and 'malfunction' is not at all problematic in this context. The difference lies merely in the goals of organisms versus those of man-made machines. Indeed, at the end of this thesis, I will offer some considerations for how the work here may bear on the questions of whether and how we could build sentient machines. This broader notion of action will help us to better think about the evolution of agency in the animal branch of life. To reiterate my previous slogan with this new vocabulary: the expansion of the organismal option-space - or as a cyberneticist might describe it the organism's degrees of freedom - are what causes a computational explosion in pathological complexity for modeller and organism alike.

Action, of course, was not invented by animals. Evaluations are found in single-celled bacteria who swim, sense, hunt, and make decisions in the broad teleonomic sense of action I employ here. Pathological complexity is a property of all life and is not restricted to animals. Yet, I do not follow the biopsychist path of using evidence for evaluations in bacteria as evidence for consciousness. Such thinking seems motivated once again by a resistance to evolutionary thinking about consciousness as something that gradually 'emerges'. It is a "category mistake", as Ginsburg and Jablonka (2019), put it, to identify consciousness with cognition in all living systems (p. 460). But we should also not commit the mistake of LeDoux that the existence of evaluation in bacteria implies that consciousness could not have evolved to aid organisms in evaluation. Consciousness should be seen as an adaptive design-solution to a computational explosion in pathological complexity that the animal branch of life was faced with, not something that exists even at the most minimal kinds of that complexity, or only at a level of human complexity.

As I've previously alluded to, Godfrey-Smith's *Metazoa* shows a notable shift towards taking the bodily challenge by Keijzer more seriously. Here, he describes action as having to be reinvented at a larger scale with new forms of coordination (2020b, p. 53). What LeDoux failed to recognize, is that when evolution has to reinvent or discover something at a new level of biological organization, things can take a very different shape and become vastly more complex. The challenge of organizing a multi-cellular unit is vastly more difficult than the challenge of organizing a single cell, even if the outcome appears superficially similar. But once this challenge has been mastered at a new level of organization, a vast possibility space has been opened for new ways of life. Multicellular action "involves coordination across vast scales from a cell's point of view" (Godfrey-Smith 2020b, p. 53). The origins of this invention in the animal tree of life can be seen in the ancestors of modern *Alcyonacea*, or soft corals, who largely stuck with *minimal action* in the form of a grasping behaviour. Here, I should reply to the same kind of question Godfrey-Smith responds to when he asks why we should emphasize actions of movement over the production of chemicals or other basic activities of life. He argues that controlled motion was a new landmark in innovation, an important evolutionary transition of action, that made the organisms objects of a new kind (Godfrey-Smith 2020b, p. 55). We can make this more precise by defining this kind of transition as the control of an organism's degrees of freedom in the service of a functional end, e.g. feeding or moving in one direction over another. Rather than conducting multiple independent actions at the same time, this transition of agency allows for whole-body actions to

the exclusion of others, and thus leading to the evolution of choice.

In a materialist picture, it is tempting to ask whether the function of the first nervous systems may reveal the origins and function of mind. The origins of the nervous system constitute a significant evolutionary event, with an entirely new mode of existence on a multicellular level. A problem Keijzer and Godfrey-Smith have have identified in a joint paper is the narrow consideration of few roles or functions the early nervous system could have played, with the "neural control of development and physiology" often being "sidelined or omitted in favour of an exclusive focus on behaviour" (Jékely et al. 2015, p. 1). One should note here, however, that bodily-control is often understood in too narrow a fashion, as the mere linking of behaviour with the right environment, but that would be a mistake. The biological world has few hard boundaries and we see something of a gradual transition from nervous systems playing the role of internal organization to taking a more outward-oriented role, such that action slowly emerges out of development (Godfrey-Smith 2002). This is why Spurrett (2020) notes that some activities of plants do constitute genuine behaviour, e.g. in a Venus fly trap (p. 7). Again, we are here not interested in intentional behaviour, but any kind of functional activity biological agents produce in the usage of their degrees of freedom.

Nevertheless, in a gradualist picture from discriminating development towards active agency, it would be a mistake to follow the move of some plant scientists such as Gagliano (2017, 2018) and fail to recognize that a significant transition towards agency took place in animal life. I acknowledge the force of their arguments that the striking cognitive and behavioural capacities of plants have been given too little attention[13], but the right move here is to strongly endorse a previously neglected gradualism, rather than deny important differences, which is a similar fallacy to that made by those who tie the presence of agency in all of life to sentience, in a biopsychist manner. As the work of Keijzer emphasizes, two important innovations that largely came together in the transition to a distinctive kind of animal agency are the nervous system and muscles that tied animal bodies together in new ways and allowed for a new set of adaptive capacities to be built on top.

An evolutionary view of consciousness must recognize gradations, and this applies to both the origins of subjective experience and the origins of the nervous system alike. Since subjective experience is often closely associated with the nervous system, it is useful to think about the very origins of nerves and neurons. Here, the role of muscles has been underestimated, with nerve-nets largely playing the role of controlling muscles in the service of adaptive behaviour (Keijzer 2015). In such a picture it makes sense for a rapid explosion of innovation to occur in the form of sensory organs and tools such as claws to engage with other organisms, but the actual explosion in complexity was enabled by a transition in the organization of action. It is here that we find the dawn of subjectivity prior to the dimensions of sensory experience and selfhood.

4.3.3 The Dawn of Consciousness

To demonstrate the significance and difficulty of this transition towards a distinctive kind of animal agency, one has to look no further than the time span it took from

[13]For a set of interesting findings about plants see Gagliano et al. (2016); Gagliano (2017); Calvo et al. (2020).

the origins of animal life to a distinctive animal lifestyle. The first definite fossils of animals date back to the late Ediacaran, though it has been contested whether to even call them animals, with their resemblance to odd flower-like shapes, long before the evolution of plants. *Dickinsonia*, which has been one of the paradigm animals of the Ediacaran, does not appear to have eyes or appendages that could give rise to interesting new ways of sensory-motor couplings. Godfrey-Smith (2020b) describes the biological imagination of this puzzling period as quiet and placid, with no evidence for interaction: "There are almost no signs of predation—no half-eaten individuals, no sign of the built-in weapons, offensive and defensive, that animals tend to have now" (p. 64). More importantly, however, is the striking absence of action in Spurrett's (2020) sense of degrees of freedom with alternative uses. Genuine action selection does not appear until much later.

Some change to this action-less picture began in the late Ediacaran (575 million years ago) until the beginning of the Cambrian, with a discernible transition taking place in animals. Waggoner (2003) distinguished three periods: Avalon, White Sea, and Nama. Strictly speaking, these three names were used to denote three major 'assemblages', i.e. findings of a collection of species that fossilized around the same time. But despite new data coming in, the paleontological picture of three distinct periods has largely remained. The first of these periods is the most important, since it is here that recent discussions in the field have placed a possible second explosion in animal complexity. Following comprehensive quantitative data analysis of the fossil evidence, Shen et al. (2008) argued that there was an 'Avalon Explosion' in the Ediacaran morphospace, mirroring the Cambrian explosion. However, while the White Sea showed definite signs of bilaterian bodies and more discernible actions of crawling on the seafloor, the Nama largely sees the disappearance of these larger complex and mobile animals, before they returned with a vengeance during the Cambrian. Here, it is useful to ask for possible mechanistic explanations of why one explosion failed, whereas the other succeeded.

Why such animal lifestyles failed despite gradual increases in sensory-motor capacities has puzzled Godfrey-Smith (2020b), who seeks to ground consciousness in the gradual evolution of just such capacities. Shen et al. (2008) ask but do not answer the question of what "constrained the Ediacara morphospace from further expansion or shift in the subsequent White Sea and Nama assemblages?" (p. 84). While it is unlikely that this Ediacaran extinction only had a single cause, the pathological complexity thesis offers us an elegant though admittedly speculative explanation for why one explosion eventually failed, whereas the other succeeded. The answer is the necessity of an evaluating system, which enables the efficient deployment of the increase in behavioural complexity through the gradual increase in sensory-motor capacities. Whereas organisms in the White Sea failed to deal with the computational explosion of pathological complexity caused by the rapid expansion of their degrees of freedom, and were thus confined to more stationary plant-like ways of life, the Cambrian saw the evolution of *Benthamite creatures* with hedonic valence serving as an impulse for efficient action-selection at the level of the organism. While this hypothesis is certainly speculative, one can readily visualize how - once this problem of efficient action-selection was solved - higher degrees of freedom became an opportunity rather than a problem. The design space for more complex behavioural capacities was opened up, leading to a rapid explosion in the Cambrian morphospace. Natural selection was free to 'experiment' in this

competition between Benthamite creatures.

Here, the status of my concept of pathological complexity as *the* explication of the teleonomic complexity of organisms once again becomes important. The complexity that matters for the organism is first and foremost a problem to be solved, not an adaptation in itself. The real problem - I hypothesize - that was solved during the Cambrian, but not the Avalon explosion, was an efficient way of dealing with the increase in complexity of action-selection. Since natural selection can only act upon behaviour by modifying the architecture behind decision-making mechanisms, McNamara and Houston (2009) argued that we need to combine the mechanistic research of physiologists with the adaptationist research of evolutionary biologists into an integrated study of function and mechanism. In the mathematical framework of state-based behavioural and life-history theory it is obvious that an increase of variables through higher degrees of freedom will lead to a computational explosion in the complexity of finding the best strategies. How can organisms solve this? This problem has been given far too little attention, despite the fact that more *agential* organisms have to solve a problem for themselves that natural selection usually solves 'for' other types of life, i.e. how to engage in the *right* fitness-enhancing activities. For much of life these are a given, but for animals with behavioural flexibility, there is a constant need to compare the returns and costs of various actions, opportunities, and dangers associated with both internal and external changes. The reason that I suspect the Avalon explosion 'failed' is because these organisms did not come up with a design solution to pay off for this complex investment into behavioural flexibility.

The animals that came closest to the animal-like lifestyles that we see today (such as *Dickinsonia* and *Spriggina*) were likely to have been some sort of grazing animals with few degrees of freedom, that moved slowly across a surface abundant in nutrients. As this resource ran out (over millions of years) these animals were no longer able to sustain their way of life (Godfrey-Smith 2020b, p. 68). Thinking about Cambrian animals, it would have seemed sensible for these animals to turn to each other as sources of food, i.e. to evolve the kind of interaction characteristic of the Cambrian explosion. Yet, for this to happen, an efficient way of organizing degrees of freedom was required, which these animals simply lacked. They did not have a valence-system with which to make behavioural complexity manageable within a decision-making bottleneck that resembles that of natural selection both for propagules and species. As Charles Sherrington (1906) argued early on in his work on the goal-directedness of the nervous system, organisms require some form of informational bottlenecking - what he called a *final common path* - in order to deal with the problem of coordinating competing actions. It is precisely here that hedonic valence makes sense as an efficient trade-off mechanism to enable adaptive behaviour (see also Cabanac et al. 2009). This in turn enabled the evolution of richer behavioural and sensory capacities, with valence serving as a common currency for evaluation. As McFarland and Sibly once nicely put it:

> [D]ecisions among different courses of action must be made in terms of a common currency, and weighted among a common set of criteria. The necessity for comparing the merits of different courses of action implies that there must be some 'trade-off' mechanism built into the motivational control system. Since the trade-off process must take into account all relevant motivational variables, it is clear that the mechanism

> responsible must be located at a point of convergence in the motivational organization.
>
> – McFarland and Sibly (1975, p. 290)

Now, it is probably too much to demand that everything we've called 'action' goes through a single bottleneck; that kind of thinking takes us back to an older Cartesian materialist model of the mind - one with a homunculus and a Cartesian theater - that Dennett sought to dispel. But Spurrett (2020) is right to insist that there is something of an intermediate position here, "a useful corrective to the tradition [...] that regards almost any convergence in a control system as a symptom of allegiance to muddled models of intelligence and cognition" (p. 11). Spurrett is referring here to the likes of Brooks (1991) who argued that representational higher-order processing for actions would lead to significant bottlenecks, with delays or even paralysis and that for this reason the world itself can serve as its own best representation (see also Clark 1997).

However, some bottlenecking is required for an animal as opposed to an pant lifestyle, since information about both internal and external states is at least to some extent opaque, and the execution of one action over another requires the combination of a variety of capacities that, as Spurrett (2020) rightly notes, itself includes trade-offs "between other possible allocations of individual capacities and combinations of them, over and above whatever the metabolic and other direct costs of this or that action might be" (p. 11). But this increase in complexity as a problem that has to be dealt with has unfortunately not received much attention. This is especially surprising because it explains why plants despite themselves dealing with complex trade-offs do not have sentience. Once organisms expand their behavioural option-space, there are significantly more trade-offs due to the sheer availability of alternative actions. There is a need for centralized control in order to efficiently coordinate the degrees of freedom of such an organism that is becoming recognizably more agent-like. Indeed, greater pathological complexity can plausibly be handled by becoming either more centralized or more modular. Since the adaptive processes of plants are modular and thus able to occur simultaneously there is simply no adaptive benefit to invest into a hedonic common currency of decision-making to exclude one action over another. Organisms with centralized organization, however, have an immediate problem of action-control to overcome that is vastly more complex than is typically appreciated in the literature and it is here that I argue sentience gradually evolved. This does not mean that the activities of plants aren't complex or don't involve trade-offs, but that there is a distinctive jump in complexity due to emergence of distinctive animal-like actions that made sentience worth having.

In debates on the evolution of consciousness even in the Cambrian, action production is unfortunately largely taken for granted (e.g. in Ginsburg and Jablonka, Godfrey-Smith, Feinberg and Mallatt), but it cannot be disassociated from action selection. In dealing with this significant transition in teleonomic complexity, a new level emerges in the Darwinian hierarchy of levels, with actions becoming distinguishable into adaptive and pathological behaviour. In order to deal with this new Darwinian challenge, organismal decisions will necessarily be made (or rather filtered/narrowed) through a number of different sub-agencies for computational reasons, but much work in neuroeconomics strongly supports the idea that there is in fact something like a global common currency for a huge variety of choice types if

not, indeed, all (Spurrett 2020).[14] It is thus not surprising that Shizgal and Conover (1996) maintained that orderly choice or - as economists would put it - revealed preferences are indicative that there must be some form of value ranking on a common scale. And valence plausibly constitutes an ancient solution to this problem in the Cambrian, which then enabled the evolution of richer kinds of felt sensory representations, such as Denton's (2006) primordial emotions like thirst and hunger directly tied to an evaluative system of efficient decision-making. An evolutionary perspective turns on its head the common view that affect is something that came on top of sensory consciousness, by making hedonic valence its most ancient capacity. The pathological complexity thesis does therefore not just imply a metaphorical sense of the existence of a common currency, but a psychologically real *felt* common currency. And it is because of this that my account offers an elegant answer to the challenge of why some sensory processes are felt and others aren't.

Unfortunately, comparative neuroeconomics remains an incredibly small field, with much of its research focused on standard model organisms such as monkeys, rodents, and birds. But what little research has been done strongly supports the idea that analogues to the human final common path are found across a wide range of the animal branch of life. This is precisely where we'd expect the presence of a common currency, in the form of specialized neural circuits designed for the exclusion of mechanically incompatible actions that can compete against each other in a preference ordering and also, as Spurrett (2020) emphasizes, "the last place [preferences] can do so" (p. 22). It is now clear that almost all vertebrates share an evaluative neural system for reward and 'punishment', with dopamine and other valence-related molecules such as serotonin sharing a deep evolutionary origin that is plausibly very ancient, rather than having been invented multiple times. The observation that some species taken to lack sentience, such as nematodes, use dopamine (alongside other such molecules implied in valance) to organize action and motor activity (see Barron et al. 2010) is only further evidence of an ancient origin of valence to help organisms achieve efficient action selection. As Spurrett (2020) notes, it is highly likely that the first implementations of preferences "were elaborations of motor control systems shared with creatures that couldn't learn, but could move" (p. 23). Indeed, it appears that dopamine is at least as old - if not older - than the invention of the bilaterian bodyplan with symmetric halves, that enabled a vast possibility increase for animal action (Caveney et al. 2006).

In the theory defended here we might thus see a valence system as the revolution needed to make animal agency 'pay off', providing an efficient action selection mechanism as the final behavioural common path in metazoans for the prioritization of some actions over others as the complexity increases through more degrees of freedom. What evidence has been gathered in invertebrates on evaluation and motivational trade-offs is highly suggestive that this is not a unique vertebrate trait. If not for the standard methodological practice in comparative cognition to only attribute these capacities to animals for which these capacities have been explicitly demonstrated, they would probably already be seen as much more basic. We are unfortunately here faced by something of a methodological artefact, in the observation that studies of motivational trade-offs (such as those of Cabanac and Elwood) have until recently been very rare in animals distantly related to us; though I will

[14]See Levy and Glimcher (2012, 2016); Pearson et al. (2014) for excellent reviews of the neuroeconomics literature.

discuss promising recent research on bees in the next chapter.

Unlike Spurrett, who tries to bracket away consciousness and emotions in his work on the origins of preferences as representational capacities, I see them as being initially instantiated through a hedonic valence-system that later became more representational by acquiring the richer sensory and integrative capacities discussed in the previous chapter on the dimensions of unity and sensory experience. These dimensions of consciousness are evolutionary add-ons, plausibly appearing in roughly the reverse order in which I have shelved them away. Indeed, it is this feature of a valence-first view that allows it to stand up to the challenge of Lloyd Morgan's canon, as it is something substantially simpler than the idea of meta-cognitive representationalist value-rankings. Cognitive complexity or 'intelligence' should not be seen as a prerequisite for simpler affective capacities (see also Browning 2019). As Dawkins elegantly puts it:

> [Y]ou don't need to be very clever to feel pain or hunger or fear. Negative emotions that we refer to as suffering are not particularly intellectual. Indeed, we often have difficulty trying to subdue our emotions with our more rational thoughts. Is it therefore not possible that the evolutionary origins of consciousness are far older than those of more refined cognitive abilities?
>
> – Marian Dawkins (2001, pp. S27-28)

Affective decision-making is plausibly more ancient than the cognitively-demanding sensory-motor couplings that evolved in the course of the evolutionary arms races of interaction. However, that doesn't mean that there isn't a design challenge here. After all, even if animals have something like subjective utilities, that doesn't mean that they will correspond to the fitness of the organism, even if this link has often been asserted (see Okasha 2018). Establishing this link is not an easy task, as Rolls (1999) elegantly points out:

> Evolution must set the magnitudes of each of the different reward systems so that each will be chosen for action in such a way as to maximize overall fitness. Food reward must be chosen as the aim for action if some nutrient depletion is present, but water reward as a target for action must be selected if current water depletion poses a greater threat to fitness than does the current degree of food depletion. This indicates that for a competitive selection process for rewards, each reward must be carefully calibrated in evolution to have the right common currency in the selection process.
>
> – Edmund T. Rolls (1999, p. 267)

Rolls and Damasio maintain that there are good reasons for thinking that these values will be updated through experience via reinforcement learning. Yet, we could question this idea. Is it really plausible that reinforcement learning can solve the complex state-based behavioural and life-history calculations that organisms are faced with? To a first approximation, we should reply with a confident yes. There is no need for subjective values to perfectly match the biological values an organism is faced with, as long as there is a sufficient degree of correspondence that

enables organisms with hedonic valence to outcompete those without such a form of evaluation. Risk, in the economic sense of uncertain value returns from actions, is an inherent factor in biological life. Environments can change and cause what is often described in evolutionary medicine literature as an evolutionary mismatch (Manus 2018). If the environment changes, an action that might previously have been fitness-advancing could now be deleterious in nature. In the early Cambrian, this factor would have only been even more important.

One way of establishing a link between valence and fitness is reinforcement learning. If something like a minimal capacity for hedonic valence comes into existence, there will be an almost immediate fitness benefit to linking this capacity to reinforcement learning. This would then also require that the subjective values can be updated, requiring further gains on the sensory side of experience. Evaluative experience would quickly become enriched in order to allow more fine-grained distinctions between what is good and what is bad for the organism, thus likely bringing the sensory dimension of consciousness soon into existence after the evaluative core has been established.

Dennett (1995a) once called organisms capable of reinforcement learning *Skinnerian creatures*, but a better term - one that is less reliant on externalist modes of thinking about what happened in the transition to a distinct kind of animal agency - would be *Benthamite creatures*. These creatures were moved by some kind of 'quasi-intrinsically' motivating states of hedonic valence and thus contained the origins of qualia. Capacities for reinforcement learning are highly suggestive of preferences, implying "both sensitivity to rewards and updating behavioural dispositions in light of reward-based consequences of earlier behaviour" (Spurrett 2020, p. 23). It is thus hardly surprising that earlier evolutionists took such learning abilities as convincing evidence that these animals can feel pleasure and pain. Natural selection plausibly led to an evolved imperative "tendency to repeat certain actions because they feel 'good' (are positively reinforcing) and this feeling good or pleasure then guides the subsequent behaviour" of the agent (Dawkins 2001, p. S23).[15] But because of the apparent ubiquity of this ability in the animal branch of life including cephalopods, crustaceans, and insects (Perry et al. 2013)[16], many have come to endorse the view that this would make consciousness too simple; that it requires something more.

But why should we hold this position? A multicellular organism that is reliably able to produce adaptive behaviour through reinforcement learning is not at all just a simple machine. The problem, as I see it, is rather how subjective evaluations can be mapped onto the actual fitness values prior to such an updating capacity evolving in Cambrian organisms. The answer, however, is not to think that Cambrian species were equipped with fixed and uniform utility functions over their sets of actions, but rather that there was great diversity even within a species that allowed species to faster exhaust the design-space of hitting upon adaptive actions. This is why the Cambrian explosion showed such a fast expansion in diversity: there were immediate benefits to hitting upon more efficient means of action selection. Natural selection simply filtered out those individuals whose subjective evaluations had a worse fit to their 'hidden' fitness values. A new realm of pathology was born, where minds could be described as pathological in virtue of how well (or poorly) they establish

[15]See also Rolls (1999).

[16]Note that all these animal groups are now slowly entering the accepted realm of animals with sentience, and I will discuss their phenomenological complexity in the next chapter.

continuity with the biological reality of the organism - which is why I say that the pathological complexity thesis is a life-mind continuity thesis. New forms of learning quickly arise in this context to create a close fit between mind and world and to ensure the organisms' Darwinian success, but I don't see a very basic capacity of hedonic valence as something that at all requires complex capacities. Rather, it is a evolutionary transition enabler upon which new capacities can be tested, to make use of high degrees of freedom in the service of an organism's ends. It is the beginning of the organism itself becoming engaged in doing life-history calculations within its own life-time. At the very dawn of this phenomenon a mere + and - feel may not have accomplished this too well, but it opened the pathway for richer forms of evaluative updating to evolve, that would solve this Darwinian challenge for organisms within their own decision-making.

The evidential burden to attribute consciousness appears to be raised as soon as we learn more about the complex capacities of other animals that we previously assumed to not be conscious. But that is a flawed approach. Firstly, a gradualist evolutionary perspective of the evolution of consciousness ought precisely to endorse a very humble origin for the origin of sentience, and secondly, the ability for affective decision-making and learning is far more complex than is typically given credit, with cyberneticists struggling to design robots achieving even the most basic successes of simple animal life. No robot has of yet been created that would be able to handle the pathological complexity of the life-histories exhibited in even the most basic of the distinctively animal lifestyles. Their failure is akin to the very same challenge Avalonian and White Sea animal agents failed to overcome.

What we see at the verge of the Cambrian explosion is the origin of Benthamite creatures with hedonic experience upon which more complex representational capacities, such as interoception, could be built. Here, I am not implying that all new capacities must go through the bottleneck of the evaluative system and be consciously experienced. But it is within the context of such evaluative agency that subjective experience *makes sense*, and plays a distinctive role for the functional deployment of the degrees of freedom a flexible animal lifestyle offers. This approach substantially narrows the explanatory gap by breaking the phenomenological complexity of consciousness down to its minimal evaluative core. Instead of a strange, apparently unnecessary add-on, subjective experience is reconceived as something almost necessary to allow an evolutionary transition of multicellular organisms from 'mere' objects subject to the whims of external forces into genuine agents/subjects, and thus makes great progress in the completion of the Darwinian revolution.

4.4 Conclusion and Further Objections

The goal of this chapter was to shine light on the close connection between consciousness and pathological complexity. The ethologists long emphasized that we need to understand organisms as teleonomic agents with life-history strategies in their natural environments. Without an understanding of what these organisms evolved *to do* it will be impossible to distinguish the normal from the pathological. Yet, this is precisely the bottleneck through which evolutionary theorizing enables us to make progress on the scientific puzzles of life. From an evolutionary point of view, health has to be understood as a measure of how an organism deals with the pathological complexity it is faced with: it is the ultimate teleonomic measure of

organismal complexity. And pathological complexity can be operationalized as the complexity of the number of parameters and constraints in the optimization problem studied by state-dependent or state-based behavioural and life-history theory. With the evolution of behavioural flexibility, the Cambrian explosion brought forth an explosion in pathological complexity of how to control action - a problem that I argued was dealt with through the evolution of hedonic valence. This in turn plausibly enabled the evolution of sensory consciousness as an enrichment in representation, providing the discriminatory richness of these Benthamite creatures, with valence serving the role of a proximate common currency. Similarly, self-consciousness plausibly evolved gradually out of further enrichments of these sensory capacities, distinguishing between self and other. Unlike strongly externalist and internalist views of consciousness, there does not appear to be some explanatory leftover - a feature of consciousness that we have left out. The explanatory gap has been significantly reduced by placing at the centre a dimension of consciousness that is inherently dynamic.

The strongly gradualist approach taken here should be attractive precisely because it minimizes the requirements for the most basic kind of consciousness, without making it an automatic feature of either life or a particular level of cognitive complexity. Consciousness, rather than simply appearing as a byproduct of other capacities, gradually evolved as a special evaluative mode of being in the Cambrian; what I have called the evolution of Benthamite creatures. In principle, I admit, one could take the view that the 'lights suddenly went on' in the evolution of this efficient way of responding to pathological complexity, yet what we see here is a gradual transition in animal agency with organisms becoming recognizably more experiential.[17] The evolution of Benthamite creatures gradually turned 'objects' into 'subjects' with their own needs becoming *felt imperative interests*, thus bringing us closer to Lewontin's demand to bring the Darwinian revolution to completion by paying attention to the *functional needs* of organisms. The evolution of subjectivity is seen as an efficient solution to the problem of action selection in the control of a complex multicellular body, enabled through hedonic evaluation. This design-solution unlocked a vast sphere of design space, including a distinctive subject-like animal way of life. Here, we see a new level of health and pathological complexity emerging beyond physiology and behaviour; the mental life of organisms can now itself be considered adaptive or pathological in virtue of how it promotes the fitness of the organism, thus giving rise to a new level of complexity, but also opportunity for animals in their engagement with the world. The appearance of sentient Benthamite creatures gave rise to the diverse phenomenological complexity of electrosensing platypuses, echolocating bats, infrared-sensing snakes, and the like, that we find in nature today.

Nevertheless, a fundamental objection against my approach here could come in the form of denying that there is anything like a common currency of values in the brain. While there is a large interdisciplinary community of researchers defending this idea and I have presented several arguments for the existence of a hedonic common currency, that doesn't mean that there aren't any skeptics. Because the idea of a common currency is so central to the core arguments in this thesis, it will be worthwhile considering the main challenges against the view. The strongest case against the common currency thesis in neuroscience has arguably recently been

[17] Godfrey-Smith (2020a) suggests the useful metaphor of "experientialization" (p. 215).

offered in an article by Hayden and Niv (2021) with the fitting title "The Case Against Economic Values in the Orbitofrontal Cortex (or Anywhere Else in the Brain)". However, as we shall see shortly, we can buy many of their arguments without selling out on the idea of a common currency of evaluation.

In their article, Hayden and Niv (2021) admit that much research in neuroeconomics has offered support for the idea that the "activity of many neurons covaries with subjective value as estimated in specific tasks" (p. 192). But they believe that the field has made too much of this connection. Hayden and Niv (2021) are right, of course, that many researchers in this field axiomatically assert "that an internal, neural, value scale must exist in the brain" and that their job is to "find this signal" (p. 194). This assumption, they maintain, has several problems. Responding to their objections will be helpful in defending my reliance on the idea of a common currency.

Firstly, Hayden and Niv (2021) argue that it is impossible to directly measure subjective values, which can only be inferred from choice behaviour. This problem, of course, is one we have encountered before. Indeed, it is the basic problem of measuring subjective experience itself. While the precise measurement of subjective values may well be a problem for the development of an economic theory of choice that can successfully predict the behaviour of agents, that does not mean that there is no subjective common currency of value. It is an epistemic, rather than an ontological, problem.

Secondly, Hayden and Niv (2021) argue that "if we had a single neural value function we called on, values elicited by different measures would match" (p. 195). But they note this is not the case, with different measures apparently implying different subjective values (see Lichtenstein and Slovic 2006). As a result, Hayden and Niv (2021) argue that "value doesn't sit in the brain waiting to be used; rather, preference is a complex and active process that takes place at the time the decision is made" (p. 195). Yet, it should be obvious that this is not incompatible with anything I have argued. It is precisely because state-based behavioural and life history theory can flexibly shift the value-rankings of alternative actions dependent on internal or environmental changes, that we should expect subjective values to be highly flexible. Again, these findings are only a problem for the idea that neuroeconomics must find a common currency that is identical to the way economists think about utility functions. That *real* subjective evaluations can at best have partial consistency and transitivity is something we should entirely expect from Darwinian agents.

Thirdly, Hayden and Niv (2021) maintain that even if neuroeconomists can make rough inferences about, say, a monkey's subjective values in a choice task and then use these to find neural correlates of these valuations, this endeavour would only be valid if we could "identify all confounding variables and regress them out" (p. 195). Again, however, this is only a problem for an economist interested in finding immutable utility values for particular outcomes. That "confounding variables include both stimulus and outcome identity, information about the state or structure of the world, the surpriseness, informativeness, and informational value of stimuli, details of the action associated with selecting of consuming the reward, including its likelihood and vigor, and the attention and arousal engendered by the stimulus" (p. 195) should hardly be surprising when we think about Darwinian agents. Subjective hedonic values are the outcomes of a complex evaluation process with an excess of information that organisms have to deal with. That there is "no

real sense in which we can read out the [']true subjective value['] of an option" irrespective of alternative options (Hayden and Niv 2021, p. 196) should be seen as a feature, not a bug. In order to allow efficient action-selection, they could not possibly be fixed.

Fourthly, Hayden and Niv (2021) argue that it "may be impossible, even in theory, to obtain a brain measure of value that is independent of behavior" (p. 195). Again, however, this is only a problem if one seeks subjective values that are encoded in a fixed manner within the brain. As I have hopefully made clear throughout this thesis, evaluation is inherently linked to behaviour. It is here that subjective evaluations must ultimately pay off, to be functional at all. And since neuroscientists are moving ever closer to disassociating the liking and wanting pathways of the brain (Berridge 1996, 2009b,a), we are also moving closer to understanding how these interact in the lives of animals, and how value is computed in the brain. Importantly, Hayden and Niv (2021) alongside with many economists, are primarily interested in the wanting-side of value, i.e. its motivational pull, not the *hedonic* liking part. But in order to understand choice and evaluation, one cannot understand one without the other. Admittedly, some might want to raise as a challenge to the pathological complexity thesis, the possibility of disassociating these two systems. The way I have talked about hedonic valence thus far may appear to lump them together and indeed I am happy to plead guilty to this charge. As Berridge (2009a) himself notes, these two systems typically converge, and it is only in rare circumstances (such as cases of addiction) that they come apart. Indeed, addiction is largely a problem unique to humans who have shaped their environments in a way that exploits this blindspot in our Darwinian design. As Ross (2020) notes, animals such as baboons and elephants that sometimes get drunk on aged berries "are at no risk of addiction [..] because they cannot cultivate sources of low-toxicity alcohol" (p. 6). For most of the history of animals since the Cambrian, we can thus safely assume that hedonic liking and wanting have been tightly linked. What is wanted or what animals have motivation to pursue is primarily driven by what they like or for that matter do not (see also Browning 2020b, p. 46). As Ginsburg and Jablonka (2019) rightly emphasize in their account of the evolution of consciousness, the liking and wanting pathways must tightly work together in order to enable animals to learn and it is not at all clear that all the neurotransmitters involved in these processes, such as dopamine, are unique to wanting or liking (pp. 352-353). Evaluations, after all, are flexible and routinely need to be updated to ensure that trade-offs are efficiently dealt with. As I noted before, reinforcement learning likely plays a crucial role in updating subjective values to correspond to the biological fitness values of an organism's pathological complexity challenges.

It thus seems that we aren't provided with any fatal arguments against the idea of a common currency of evaluation as it is defended here. Only those attempts that try to find a complete vindication of standard economic theorizing about utility functions within neuroscience will be left disappointed, but that of course is no problem for us here; nor do I believe that the majority of neuroeconomists have such conservative ambitions. The goal is to understand the evolution of real agents and the Darwinian implementation of a common currency view may well conceptually re-engineer the typical economic picture of agency. Furthermore, Hayden and Niv (2021) even admit that a single common currency of evaluation would significantly simplify evaluation processes that involve multiple dimensions, which perfectly fits

with the pathological complexity thesis. Yet, given their objections to how common currency views of value are often formulated it is not surprising that they think that alternative choice and learning rules could equally explain these results. However, as I have shown, we may well agree with many of the criticisms Hayden and Niv (2021) raise. In the picture I defend here, not all actions are the result of hedonic impulses. Indeed, we can readily accept the point raised by LeDoux and Dawkins that most evaluative processes in the brain are going on unconsciously. What I have given is an explanation for why a particular subset of evaluations are felt, i.e. where multiple dimensions of trade-offs are engaged with each other we can expect that it pays off for hedonic valence to play a role at the final behavioural common path. My argument is not a defense of the economists' idea that all decisions rely upon such a common currency. As I noted above, Brooks (1991) is right that such a significant bottleneck would leave animals unable to make quick efficient decisions. But as so often in biological design, there is a trade-off here, including a set of circumstances where *some* bottlenecking is highly useful. And as I hope to have successfully argued here, it is a computational explosion in pathological complexity primarily driven by increased degrees of freedom that made a Benthamite mode of being worth having. We are thus provided with a functionalist answer to the demand of LeDoux (2019) that animal consciousness researchers need to provide a means of distinguishing conscious from nonconscious processes of evaluation.

With this hedonic model in place, the next chapter will try to build a firm understanding of phenomenological complexity as a response to pathological complexity, by placing the other dimensions in the context of a general theory of consciousness based on the evaluative side of experience. It will enable us to make predictions regarding the likely subjective mental states of animals based on the evolutionary history and ecological challenges faced by them *in nature*, which can then be tested through various experimental means.

Chapter 5

Pathological Complexity meets Phenomenological Complexity

> *Granting that hypotheses [about animal minds] are difficult to test by currently available procedures, the tentative consideration of their plausibility might pave the way for thoughtful ethologists to devise improved methods to study when and where animal consciousness may occur and what its content may be. The future extension and refinement of two-way communication between ethologists and the animals they study offer the prospect of developing in due course a truly experimental science of cognitive ethology.*
>
> – Donald Redfield Griffin (1981, p. 171)

5.1 Introduction

We have reached the stage at which we can synthesize the previous chapters to make progress on the cognitive ethologists' goal of assessing the phenomenological complexity of other animals. With a model in place of consciousness grounded in evaluative agency, it is time to return to the other four dimensions that we previously shelved away from the most ancient kind of subjective experience. It is here that we find perhaps the strongest virtue of the pathological complexity framework in being able to draw on modern life-history theory to make predictions regarding the subjective experience of other animals based on our ethological understanding of what it means to be an octopus, a bee, or raven in their natural healthy lifestyles.

But the goal of this chapter is not only to show that the other dimensions of consciousness can be readily explained within a theory of consciousness centered on evaluative experience. We will also put the pathological complexity thesis to the test by responding to an immediate challenge to the pathological complexity thesis raised by Godfrey-Smith (2020c,b), who argues that that there could be a phylogenetic split in conscious experience, with some animals having evaluative experience while lacking the sensory side or vice versa. This is an interesting challenge to the pathological complexity thesis, since I have not only argued that it is the evaluative side of consciousness that emerges first in evolutionary history, but also that is is precisely within the context of hedonic evaluation that explains the qualitative experience of the other dimensions. In my continuous emphasis on phenomenological

complexity, however, we should at least take seriously the "idea of deep differences between varieties of subjectivity" (Godfrey-Smith 2020b, p. 217). If we find animals who only have sensory experience without the evaluative side, as opposed to say some loss in evaluative richness, that would be an empirical challenge to the pathological complexity thesis.

To test my thesis we will first look at the gastropods (snails and slugs) and secondly the arthropods (in particular crustaceans and insects). Admittedly, only insects (hexapods) constitute a real test case, since Godfrey-Smith (2020c) uses gastropods as a potential case that "may have relevant evaluative complexity" to be experienced but lacking sufficient complexity on the sensory side (p. 1153). Indeed, my discussion of gastropods will primarily serve as evidence *for* the pathological complexity thesis: the possibility of minimal consciousness in the sense of hedonic valence, without the other dimensions of consciousness. In insects, however, Godfrey-Smith maintains that they have only simple evaluative capacities, whereas their sensory capacities are sufficiently rich to make it at least plausible that they could have sensory experience without the evaluative side. Reviewing the evidence in the literature for this view, I will ultimately reject this challenge to my thesis.

After this defense of the pathological complexity thesis, I will turn to the dimensions of unity and selfhood, and discuss them in the context of the species-specific pathological complexity of cephalopods, non-avian reptiles, fish, and birds.

Chapter Outline

This chapter is structured as follows: In Section 5.2 'Gastropods: A Sluggish Way of Life', I use the case of gastropods to support the motivation of the pathological complexity thesis to seek the origins of consciousness in evaluation. In Section 5.3 'Arthropods: A Robotic Way of Life', I respond to the challenge that insects might have sensory experience without evaluative experience. In Section 5.4 'Octopuses: A Disembodied Way of Being', I will discuss the dimension of self-consciousness through the biological lens of the octopus, who arguably provide the best biological challenge to human-centric ways of thinking about the role of the self. In Section 5.5 'Fishes and Non-Avian Reptiles: Dis-Unified Ways of Being', I will discuss these natural split-brain patients to assess the possible adaptive benefits of more disunified forms of consciousness. In Section 5.6 'Corvids: A Cunning Way of Being', I discuss the adaptive roles of diachronic unity through a comparison of corvids and those humans with aphantasia. Finally, Section 5.7 'Challenges, Conclusion, and Further Directions' will summarize this chapter, offer some responses to potential objections, and explore potential directions for the further development of the pathological complexity framework.

5.2 Gastropods: A Sluggish Way of Life

The first class of animals I shall discuss are gastropods (i.e. snails and slugs), which form a large group of invertebrates and evolved in the sea during the Cambrian, though their precise origins remain contested (Parkhaev 2007). Like the cephalopods, who are now accepted by many to possess subjective experience, gastropods belong to the phylum of molluscs, though their nervous system is generally simpler. Unlike the cephalopods - and in particular octopuses - gastropods have

received only little attention in debates on animal consciousness. Despite their taxonomic diversity, their lifestyles have been considered too slow and too uninteresting, compared to the extreme behavioural flexibility, tempo, and intelligence of their octopus relatives. One might thus be tempted to categorize their pathological complexity as too insignificant. But here we should heed the warning by Dennett (2019a) that our imagination is in many ways shaped by what Wittgenstein dubbed *Lebensform* (translated: form of life), that is "our linguistic communities, the commonalities that are apt to confound our thinking with parochiality" (p. 2). If we observe animals distantly related to us and with very different ways of life, we will be influenced by what Dennett (2019a) nicely expressed as their *speed* and *rhythm*:

> [I]f cephalopods moved in the clunky way of most existing robots, then in spite of the manifest purposiveness of their motions, it would be quite comfortable to suppose that they were some kind of zombies, marine robots with eight or ten appendages.
>
> – Daniel C. Dennett (2019a, p. 2) [emphasis in original]

Gastropods, of course, appear even slower than many sophisticated robots. In this context it is also worth pointing out that it is hard for many to avoid the illusory attribution of consciousness brought about by sophisticated ventriloquists[1] or robots designed to move in very human-like ways.[2] Care must be taken not to put too much faith in our intuitive willingness to grant consciousness to some entity. But one need not draw on thought experiments or artificial cases such as robots to get this point across. While Dennett (2019a) sees giant sea slugs of the genus *Aplysia* as nothing more but an excellent model organism due to its simple nervous system, Godfrey-Smith (2020b) replies from his personal experience that unlike other smaller slugs with similar nervous systems, "all one has to do is scale them up to giant *Aplysia* size and have them move at a gallop rather than a slow crawl, and suddenly experience in these animals seems almost inescapable, or at least far more feasible" (p. 216). What should one make of such observations? The lesson, I suggest, is that it is misguided to make our attribution of consciousness contingent upon our ability to occupy other's points of view. We are too influenced by our own distinctive human perspective.

Whether we are faced with echolocating bats, crawling gastropods, buzzing bees, or oddly moving octopuses, we have to remind ourselves that these animals live different *kinds* of lives, which ought to bear out in different kinds of minds. Their conscious experience should reflect their life-histories, which are strikingly different from ours. Perhaps humans simply lack the imagination required for this task. We are, after all, notoriously bad at putting ourselves in the shoes of others - this not only applies to strangers of a different sex, race, or origin, but even to our closest family members. More notable perhaps is our inability to even recall what it was like to be us in the past or predict what it will be like to be us in the future. The best-known example of this is the *disability paradox*: while people with

[1]The neuroscientist and ventriloquist Michael Graziano (2016) frequently uses a small Orangutan puppet named Kevin in his talks to demonstrate how quick we are to attribute consciousness to other entities.

[2]Dennett (2019a) reports of Cog, a robot developed by a team at MIT that "moved its arms and eyes and head with such humanoid vivacity and even grace that naive observers often blurted out loud their startled conviction that it was conscious" (p. 2).

disabilities often report a high quality of life, non-disabled people believe this to be a cognitive mistake - imagining much bigger losses in quality of experience when they imagine themselves to have these disabilities (Albrecht and Devlieger 1999). Introspection can be positively detrimental for the assessment of another being's subjective experience, even in our own case. Here, gastropods constitute a beautiful case for the need of an ecological bottom-up approach.

Prominently, Feinberg and Mallatt (2016) argue that evidence for consciousness in gastropods is lacking, but they also admit that there is some evidence pointing towards the affective side. Nevertheless, they end up denying consciousness to gastropods since they are said to "lack the brain complexity one would expect for consciousness" (p. 192). This, of course, raises the same validation problem we saw for *Integrated information theory*, of whether we already know what complexity would actually be required for consciousness. Godfrey-Smith (2020c) evaluates the evidence in a different way by emphasizing that gastropods may be a case for sufficient richness in evaluative capacities to have evaluative consciousness while lacking the other dimensions. If so, this would provide strong support for the pathological complexity thesis: we could have animals around us in the here and now, rather than just at the origin of consciousness in the Cambrian, with a minimal sense of hedonic evaluation without the other dimensions. A theory of consciousness based on the human case is undoubtedly prone to fail in its recognition of such 'marginal' cases, so it is useful to look at their 'lived experience' from their own point of view by using the pathological complexity framework.

Evaluative Experience

In his emphasis on the evaluative capacities of gastropods, Godfrey-Smith draws particularly on the work of Terry Walters, who has been one of the front-runners in advancing our understanding of gastropod *skills*.[3] As I've emphasized in my discussion of the evaluative dimension, we should not simplify this dimension to contain only pleasure and pain, in the sense of two highly specific mental states, but rather include any sort of subjective experience that has a positive or negative valence. This can also include medium-term and long-term states such as emotions of anger or fear, and moods such as pessimism. As we will also see in the discussion of insects that will follow, we should be open to the existence of all kinds of negatively valenced states, and not limit them to human-like cases of pain involving rich sensory representation.

Crook and Walters (2011), for instance, argues that gastropods show *nociceptive sensitization*, which Godfrey-Smith (2020c) describes as "a heightened sensitivity after damage" (p. 1155) and sees as compelling evidence for perhaps a minimal sense of evaluative experience. What this work has shown is that when gastropods are exposed to aversive stimuli such as electric shocks, they not only react to this with an immediate behavioural response, but there also appears to be a long term change in behavioural 'character'. Crook and Walters (2011) argue that *Aplysia* show a conditioned fear-like motivation state when exposed to a neutral chemosensory stimuli when it has been associated in the past with an electric shock (p. 189). Furthermore, associative learning was demonstrated in gastropods as early as 1981 (see Carew et al. 1981; Walters et al. 1981; Colwill et al. 1988), an ability which has

[3]See Walters (2018) for a recent review.

often been linked with consciousness. When the smell of a shrimp was paired with an electric shock, *Aplysia* showed surprising future responses to this stimuli, such as (i) freezing in response to the smell even in the absence of electric shocks, (ii) halting feeding when exposed to the smell, and (iii) withdrawal, escape, and defense responses when the smell was paired with light touch (Crook and Walters 2011, p. 189). Godfrey-Smith (2020c) considers this range of responses compelling evidence for a "pervasive state of negative readiness" linked to the feelings side of subjective experience (p. 1155). From exchanges with Walters, Godfrey-Smith reports that he is more cautious about attributing sentience to them, but acknowledges the striking teleonomic rationale of an "ability to maintain functional 'awareness' of injury-induced vulnerability until the vulnerability subsides (perhaps until adequate repair of damaged body parts has been achieved)" (Walters 2018, p. 13) [cited in Godfrey-Smith (2020c, p. 1155)].

If the pathological complexity thesis is right, then this is exactly how the vulnerability of complex multicellular organisms gives rise to hedonic experience. One may even see these negative mood-states as involving a minimal sense of self and diachronic unity, but these features need not be part and parcel of the subjective experience of an animal in order to associate particular stimuli with a negative valence. After all, even humans can have a negative emotional reaction to an event or food item without the ability to consciously draw the connection to a previous negative encounter. Nevertheless, it is tempting to think that episodic memory can be readily explained as something built on these capacities once they are in place, and we should resist the thought that current boundary cases for the attribution of sentience must be anything like the animals in the early evolution of subjective experience. It is not at all implausible to think that the presence of a hedonic evaluation system quickly gives rise to further increases in phenomenological complexity.

Furthermore, Godfrey-Smith (2020c) praises Walters for highlighting the ecological lifestyles of *Aplysia*, which often involve longer life-cycles of one to two years more than is common in many insects. If we try to explicate the pathological complexity of gastropods we will quickly find an additional rationale for these long-term mood states. Because their behaviour is relatively limited in comparison to that of many other animals that are discussed as potential bearers of sentience, wound-tending does not appear to be even within their behavioural option space. Yet, this doesn't mean that gastropods aren't vulnerable. Unlike insects, whose bodies are hard, many gastropods lack even shells to protect themselves. But whereas insect bodies can often not be 'repaired', hence making protection superfluous, gastropods almost constitute an opposite case, with excellent if not extreme abilities to heal. As long as wounds are not mortal, they will quickly restore their bodies to a healthy state of normalcy.

An extraordinary case in the genus *Elysia* cf. *marginata* reported by Mitoh and Yusa (2021) has recently gained a lot of attention, since these slugs have been shown to be able to decapitate their own heads from their body, which includes shedding of the whole heart, in order to rid themselves from a potentially parasite-infested body. This is an extreme case of *autotomy*, (i.e. the behavioural strategy of deliberately shedding body parts), enabled by the special regenerative modes of being of gastropods. This is one way of responding to pathological complexity and thus entirely healthy behaviour. But it is also precisely in this context - in which behaviour is limited, and bodies are vulnerable yet allow for healing - that it makes

sense to invest both in short term states of pain and in longer term mood states such as fear or pessimism. Their life cycles last long enough to make such capacities a useful investment. However, we should not make too much here of the associations with certain rich human emotions and mood states. What we are interested in are these states as natural phenomena, which makes the human case a special case rather than a typical exemplar.

Due to the small nervous system that has made *Aplysia* a model organism to begin with, these results provide compelling functional evidence that a minimal degree of sentience may be present in these slow and vulnerable creatures. This view isn't anti-neural as much as it is gradualist. Because *Aplysia* belong to the largest sea slugs, especially sea hares (*Anaspidea*) among them, such as the California sea hare (*Aplysia californica*), that are comparatively much more active - their movement resembling a "gallop rather than a slow crawl", as Godfrey-Smith notes - it can be hard not to grant them experience (2020b, p. 216). But despite their behavioural differences from smaller sea slugs who have very similar nervous systems, the seeming lack of intuitive draw towards attribution of sentience in the latter group may merely be a matter of perspective, with Godfrey-Smith arguing that once the smaller relatives are scaled up to the largest among the Aplysia, it becomes difficult to draw a hard boundary of experience; doubly so if their movement is sped up. A gradualist picture is tempting here, and fits better with the actual data than the demand for a hard line. Even tiny slugs and their ancestors may possess an evaluative common currency of valence to deal with motivational trade-offs, despite a lack of capacities in the other dimensions. But to assess this idea, let us also examine another dimension, in order to determine whether or not they really do have only evaluative experiences.

Sensory Experience

In the previous section I mentioned that gastropods, when compared to insects, seem to have fewer degrees of freedom in their behavioural repertoires. Furthermore, they have much simpler sensory capacities, though there are some exceptions. Godfrey-Smith (2020c), for instance, notes that sea elephants or heteropods (*Pterotracheoidea*) have something of a borderline case of type IV eyes, which might provide compelling evidence for sensory experience on the visual side. What also distinguishes the lifestyles of these species is that they are much more mobile - they have fins for free swimming and engage in predation, in contrast to most gastropods that live on the ground. The pathological complexity they are faced with is therefore different from the usual sluggish gastropod way of life.

For these swimming gastropods, with lifestyles more closely resembling the pathological complexity of fish and cuttlefish, we can make predictions regarding the likely richness in their sensory capacities. If sensory capacities are found in various degrees of complexity within a branch of life that is already a likely contender for minimal sentience, the pathological complexity thesis appears to gain striking support for the close relationship between lifestyle and experience. Most gastropods, however, appear to only have a "sliver of the features that make for experience in us" (Godfrey-Smith 2020b, p. 262), and this sliver appears to be mostly on the evaluative side, providing compelling evidence for the independent existence of evaluative experience without strongly representationalist sensory capacities (or, for that matter, the other three dimensions). It is plausible that a transformation of this

largely evaluative mode of subjective experience towards a richer representationalist form would simply not pay off for gastropods with their unique life-histories, but that should not be taken as evidence that they do not have consciousness at all. As Ginsburg and Jablonka (2019) note, "in gastropod mollusks that have more complex brain ganglia than the sea slugs, the presence of elementary mental representation is an open question" (p. 394), but this will have to be investigated further.

Given the evaluative richness of gastropods it is not at all implausible that they could also have a low degree of sensory experience. Godfrey-Smith (2020b) himself admits that gastropods appear to have a rich sense of smell (p. 215). In thinking about this dimension of consciousness, we appear to often be all too tempted to give priority to vision, which is precisely why I have replaced the term 'perceptual richness' with 'sensory richness' in my discussion of Birch et al. (2020). Nevertheless, since agreeing with Godfrey-Smith here is not detrimental to the pathological complexity thesis, let us now turn to the case of insects, which provides a stronger challenge. If Godfrey-Smith is right that insects possess sensory without evaluative consciousness, this could undermine my thesis altogether.

5.3 Arthropods: A Robotic Way of Life

Whereas Godfrey-Smith (2020c)'s arguments for the presence in gastropods of evaluative experience without the sensory side provide strong support for the pathological complexity thesis, his arguments for the existence in insects of sensory experience without the evaluative side provides an interesting challenge that we will have to overcome. Godfrey-Smith (2020c) suggests that complexity in sensory "capacities might be understood as involving complexity in discrimination or in downstream processing" (p. 1153), but emphasizes the latter as being more important for considerations of subjective experience. While Birch et al. (2020) emphasized discrimination in sensory richness, Godfrey-Smith's emphasis on downstream processing is certainly reasonable due to a recognition of how many discrimination-activities the brain is engaged in without subjective experience, even in humans. For the purposes of the discussion here, I remain neutral on the question of which side matters more, since the pathological complexity thesis sees sensory experience as something operating within an evaluative sphere. Let us therefore look at the possibility that sensory experience could exist without such an evaluative space in which different sensory stimuli are being evaluated against each other.

Insects are part of the arthropod branch of life and constitute the great majority of arthropod species (in addition to that of all animals). They are estimated to have originated only roughly 479 million years ago during the early Ordovician, which suggests - as Misof et al. (2014) point out in a landmark study in Science - that they have evolved in response to the plants which started to colonize the planet around the same time (see also Labandeira 2006). However, the arthropod group, which also includes crustaceans (e.g. crabs, lobsters, and krill), arachnids (e.g. spiders and ticks), and myriapods (e.g. centipedes), are a much earlier Cambrian invention - indeed, they constitute the paradigm phylum of the Cambrian explosion, leading the way for a special animal way of life. Their name, being a conjunction of the Ancient Greek words for 'joint' and 'foot', is a fitting description for a mode of being consisting of hard shells, multiple segments, and typically many appendages (Budd and Telford 2009), that nevertheless shares a common active animal lifestyle with the

'soft' and 'sluggish' gastropods. But despite sharing a high degree of pathological complexity, it plays out differently in both groups and this might make it tempting to think that arthropods could evolve sensory consciousness without the presence of evaluation. To examine this further, this time we will begin with the sensory side of things.

Sensory Experience

Unlike the soft-bodied gastropods, arthropods seemingly overflowed in the Cambrian, with trilobites making up much of the fossil record. Partially, this is due to their possession of an exoskeleton, which simply fossilizes better, but their presence emphasizes much of the change that took place during the Cambrian. An exoskeleton makes *sense* as a protective shell against others, with appendages such as feelers and claws clearly existing in response to other subjects, whether prey, partner, or predator. Godfrey-Smith (2020b) describes arthropods as having "invented a new way of being an animal, with a skeleton that scaffolds and organizes complex actions. They also invented claws, and to go with them, image-forming eyes" (p. 80). The life-histories of arthropods involve frequent interactions with others, so it is unsurprising that their sensory capacities and behavioural repertoire surpasses that of gastropods - a richness that is also emphasized by Feinberg and Mallatt (2016). Indeed, if we think about the pathological complexity of insect life, it becomes surprising that few have granted them a minimal sense of subjective experience despite the prevalent vision-centric model of consciousness, since many insects have been shown to have sophisticated sensory capacities, especially high-resolution vision.

Notably, Godfrey-Smith (2020c) focuses on the much-studied bees and fruit flies (*Drosophila*), since it is here that we can examine flight as a complex behaviour that "involves dealing with complex spatial layouts and making self/other distinctions with respect to the causes of sensory events" (p. 1153). Indeed, in the ecological framework that I try to build here, flight constitutes the paradigm case for an explosion in pathological complexity, due to the rapid feedback between sensing and action. Godfrey-Smith (2020b) sees the sensory processing of flying insects as a plausible candidate for subjective experience, since it is such "feedback that contributes to a point of view" (p. 208). That flying creates a new challenge of complexity is not a new idea. In his very first publication, the British evolutionary biologist and former aeronautical engineer John Maynard Smith (1952) already emphasized the importance of a sophisticated nervous system in the evolution of flight, for both birds and insects. He argued that the evolutionary origins of flight must have required flight stability via a long tail, since early animals lacked the sensory richness and nervous system complexity to control such a flying body - similar to how pilots require a stable plane in order to be able to fly it. While such a tail lowers maneuverability, it greatly increases flight stability. Yet, Maynard Smith (1952) argued that "in the birds and at least some insects, and probably in the later pterosaurs, the evolution of the sensory and nervous systems rendered the stability found in earlier forms no longer necessary" (p. 129). The evolutionary advantages of unstable flight, he argued, would be the ability to turn more rapidly in the air and to be able to fly at slower speeds without falling (p. 128). Taking a design stance toward the problem of flight makes it obvious how rich the complexity of this problem really is. Free fall can mean death. But Maynard Smith made these comments in relation to birds, and insects are a different case.

Because insects are so small, air resistance will stop them from gaining enough fall speed to cause serious injury. Nevertheless, it is precisely because of their size that it is more important to focus on the organization of the insect nervous system and its decision-making speed, rather than the mere number of neurons. Broom (2014), for instance, notes that the critical flicker fusion tests have shown blowflies to have the fastest temporal processing speed, "perhaps not a surprise to all who have tried to catch one" and that they may "perceive humans as slow and inconsequential" (p. 118). Such findings could not only suggest that insects possess the sensory side of subjective experience, but also that it might be richer than in larger animals, including us - a conclusion we might even want to extend to their experience of time. Indeed, there is plenty of evidence that the more maneuverable and fast fly (or for that matter, bird) species are, the richer their temporal resolutions (Laughlin and Weckström 1993; Potier et al. 2020), thus supporting the link between pathological and phenomenological complexity.

Yet, flying is in a sense a very new lifestyle and from an evolutionary perspective we should think of flying insects as a distinct category, just as we treated the giant *Aplysia* as non-representative for the entire gastropod class. Their pathological complexity will be much higher due to their different lifestyle and in many ways resemble that of flying birds, which should lead us to engage in comparative thinking about the convergent evolution of phenomenological complexity. Here, I would also like to point out that we should give evolutionary thinking some priority over taxonomic thinking, which has historically been used mostly to claim that consciousness wouldn't exist in animals belonging to certain parts of the tree of life. Nevertheless, flying insects are themselves an evolutionary product and it is unlikely that their non-flying ancestors had poor sensory richness, since it is probable that at least a certain threshold of such complexity would be required. Though, following Maynard Smith, it could also be conceivable that the ancestors of flying insects had a long tail that provided a useful evolutionary scaffold for the evolution of flight, one that could later be discarded with the evolution greater sensory capacities.[4] So while it makes evolutionary sense to attribute rich sensory consciousness to these insects, we should be careful not to necessarily attribute this capacity to all insects.

My aim here is importantly not to prove that insects have sensory experience, only that it makes evolutionary sense within the context of the pathological complexity framework as an enrichment of evaluative capacities that became more representational. In order to address Godfrey-Smith's challenge that the sensory side could exist without the evaluative side, we will now simply assume that these animals are conscious, in order to evaluate whether such a separation makes sense for a conscious animal.

Evaluative Experience

One of the main reasons we should not think of sensory experience existing without evaluative experience, is that it is ultimately through evaluation that richness on the sensory side 'pays off'. As Broom (2014) rightly acknowledges, the flicker fusion tests are closely related to the ability of animals to evaluate and make decisions, which is why he describes blowflies as having the "fastest evaluation" (p. 40). Yet,

[4]See Caporael et al. (2014); Griesemer (2014); Veit (2022e) for discussions of 'scaffolding' in evolution.

Godfrey-Smith does think that there is a compelling case to be made for a separation between the evaluative and sensory sides of experience in insects.

In particular, Godfrey-Smith draws on an old but influential mini-review by Eisemann et al. (1984) in order to establish that "all known insects appear completely unconcerned about even severe body damage, despite having nociceptors. Wound-tending has never been seen in an insect, and after injury these animals just continue, as best they can, with the behavior appropriate to the circumstances" (Godfrey-Smith 2020c, pp. 1153-1154). But this is partially a misrepresentation of even this early work on the possibility of insect pain. Indeed, Eisemann et al. (1984) cite early experimental work by Hentschel et al. (1982) that showed grooming activity in response to damage, and presented it as something to be *explained*. Eisemann et al. (1984) explicitly recognize an "increase in both general grooming activity and specific grooming of a wound site observed after experimental puncturing of the abdominal wall of the cockroach *Periplaneta americana* (L.)" (p. 166). While it is true that Eisemann et al. (1984) argue that insects do not feel pain, they do so in a very measured way, only stating that "the evidence from consideration of the adaptive role of pain, the neural organisation of insects and observations of their behavior does not appear to support the occurrence in insects of a pain state, such as occurs in humans" (p. 167). That they see the evidence as far from conclusive is also emphasized by their call to endorse Wigglesworth's earlier "recommendation that insects have their nervous systems inactivated prior to traumatizing manipulation. This procedure not only facilitates handling, but also guards against the remaining possibility of pain infliction and, equally important, helps to preserve in the experimenter an appropriately respectful attitude towards living organisms whose physiology, though different, and perhaps simpler than our own, is as yet far from completely understood" (p. 167).[5]

This then makes it somewhat puzzling as to why Godfrey-Smith (2020b) similarly repeats his assertion in *Metazoa* that "insects have still never been observed tending and grooming injuries; that claim from the old no-pain paper still holds up" (pp. 211-212).[6] Just because insects have not been shown to engage in sophisticated "protective behavior towards injured body parts, such as by limping after leg injury or declining to feed or mate because of general abdominal injuries" (Eisemann et al. 1984, p. 166) does not mean that no grooming-like behaviour has been observed - *even if* it could be explained in a way unrelated to pain. The way Eisemann et al. (1984) deal with Hentschel's observations is to point out the "contra-adaptiveness of this response in relation to wound healing" (p. 166). But we have to distinguish the adaptive value of such behaviour from its support for the presence of subjective experience. It may very well be the case that not all grooming behaviour is adaptive, no less so than the scratching of human wounds is. Pain could be invoked as a cause as long as a general experienced negative valence exists regarding damage or potential damage. Indeed, this might even be seen as supporting the presence of negative valence as opposed to a mere 'mechanical' response.

My argument here, however, should not be read as me endorsing the presence of pain in insects - this is yet to be established. I only argue that the case is not as straightforward in insects as Godfrey-Smith makes it seem. Nevertheless,

[5]See also Wigglesworth (1980).

[6]In personal communication Godfrey-Smith admitted that he should not have used the term 'grooming' in his list.

it is certainly true that insects - more so than perhaps any other complex agent-like animal group - have been observed to be apparently oblivious to all kinds of bodily damage and injuries. They engage in sex and feeding while being devoured, soldier on despite damages, and even eat their own insides when they are leaving behind their body due to damage (Eisemann et al. 1984; Adamo 2016; Walters 2018). In order to understand whether or not such behaviour is functional, we will have to understand the pathological complexity faced by insects. Godfrey-Smith (2020c) notes the "ecology of insects is also relevant" (p. 1154), but for a true cognitive ethology it should be our primary source of information when extending the Darwinian revolution to consciousness. The references by Godfrey-Smith (2020c) to the life-histories of insects vs crustaceans are particularly interesting here.

Whereas most crustaceans live in the water, leading similar lifestyles to their Cambrian ancestors, insects have predominantly branched towards a life on land.[7] Yet, whereas Godfrey-Smith wants to deny evaluative experience in insects, he grants it to crustaceans, where wound-tending has been shown (Birch et al. 2021). The work of Elwood and his collaborators (Appel and Elwood 2009; Elwood et al. 2012) has been particular influential in this context, which - similarly to the early work by Cabanac - studied the evaluative trade-offs decapod crustaceans (shrimps, crabs, and the like) are engaged in. For instance, hermit crabs have shown that they are making state-based decisions on whether or not to leave their shell when receiving electric shocks, dependent upon both the predicted presence of predators and the shell value. In a review for the Department for Environment Food and Rural Affairs (Defra) of the UK that covered a vast range of further evidence - so vast in fact that it could fill an entire chapter - Birch et al. (2021) recently argued that decapod crustaceans are sentient. This has led to them being added (alongside cephalopod molluscs) to the list of animals included in UK animal welfare legislation. So it appears plausible that the evolutionary ancestors of these animals in the Cambrian already possessed sentience.

If we grant this, however, then this admittedly transforms Godfrey-Smith's challenge to the pathological complexity thesis. Instead of sensory experience arising independently in its own right, the challenge now appears to be explaining a reduction of richness on the evaluative side in insects once the sensory side has come to play a more important role. After all, these results are convincing enough to have motivated Tye (2016) to call his book *Tense Bees and Shell-Shocked Crabs: Are Animals Conscious?*. Allen and Bekoff once criticized as anti-neural Griffin's suggestion that bees might have more of a use for subjective experience *because* their nervous systems are so small (1997, p. 153), but I take it that Griffin's motivation was to highlight the important insights from the ethologists to emphasize the lifestyles of animals prior to conducting laboratory experiments. More so, we may be able to explain the observations of Godfrey-Smith between different levels of richness within the pathological complexity framework, without having to give up on the idea that hedonic valence is required for and lies at the core of qualitative experience.

Within the peculiar pressures of life on land, most insects have evolved very short and routinized life-histories that differ from the comparatively longer and "less regimented lives of their marine relatives studied by Elwood" (Godfrey-Smith 2020c, p. 1154). While there are exceptions to this rule (Maruzzo and Bortolin 2013; Suzuki et al. 2019), insect limbs generally do not regrow and there is little evidence

[7] Crustaceans are likely a paraphyletic group (Blackstone 2001).

that there is adaptive value for them in protecting injuries, since their capacities for healing are quite poor. Godfrey-Smith (2020c) describes this lifestyle as being about *soldiering on* even in the face of pathologies (p. 1154). Now, this makes sense in a semelparous lifestyle and especially in eusocial insects, where individuals are replaceable. One might expect bees or ants to have sophisticated sensory capacities for finding food sources but to be less rich on the sensory side *in order to* focus on their tasks. But does this really show that the evaluative side has been lost?

Godfrey-Smith (2020c) admits that bees have been shown to avoid noxious stimuli such as heat, but notes that this could be a mere reflex, not necessarily involving subjective experience. A compelling line of evidence in this context are various forms of learning ability, since they are commonly taken to constitute convincing evidence for the possibility of evaluative experience. Unfortunately, the abilities of insects to learn are routinely underestimated by both scientists and philosophers. Figdor (2020), for instance, criticizes Dennett's (1973) discussion of the Sphex wasp - "characterized as an inflexible machine incapable of any learning" as a clear case of "confirmation bias" (p. 2). But there is no longer a debate on whether insects are capable of reinforcement learning (Allen et al. 2005; Elwood et al. 2012), and insect consciousness is starting to be taken as a serious possibility. Which forms of learning constitute the best kinds of evidence for consciousness, however, remains contested. Ginsburg and Jablonka (2019) provide a good overview of this debate and argue that there is unlimited associative learning that provides something like a transition marker that animals are conscious; though they do not mean to say that the absence of unlimited associative learning necessarily shows that consciousness isn't present. I am very skeptical that we can actually find anything like a definite marker, since consciousness can come in a diversity of forms, but that is not an objection to the idea that sophisticated forms of associative learning constitute good indicators at least for a certain richness or even transition of consciousness, as opposed to its presence. But as with gastropods, we should also highlight evidence for nociceptive sensitization as indicative of evaluative richness in insects - which is notably also highlighted by Tye (2016). One peculiar result that Godfrey-Smith (2020c) emphasizes, is the presence of sensitization in *Drosophila* larvae, as opposed to its later life-stages (p. 1156). Too much focus, he notes, might have been given by Eisemann et al. (1984) and Groening et al. (2017) to the absence of pain in adults:

> Another factor in insects not highlighted so far, one related to life on land, is the differences between larval and adult states. Many insects lead two lives, in effect, one on each side of a metamorphic divide, with extensive breakdown and reconstruction at that stage. In the kinds of insects considered here, it is the adult who has acute sensing that controls complex motion; the larva does not.
>
> – Peter Godfrey-Smith (2020c, p. 1156)

Drawing on Sprecher et al. (2011), Godfrey-Smith (2020c) emphasizes that larvae have very simple eyes - in *Drosophila* only the small number of 12 photoreceptor neurons dedicated for vision, much simpler than in the adult stages. Yet, in contrast to the apparent obliviousness to damage in adult insects, Godfrey-Smith (2020c) also highlights the work of Walters (2018) that showed larval stages of *Manduca* and *Drosophila* to have nociceptors and nociceptive sensitization. What we find in

insects is a striking disconnect between the pathological complexity faced by the larval and the adult stages.

Partially, such observations are due to the extreme diversity of insect life cycles, which explains the presence of a huge variety of alternative life-history strategies; including such odd examples as male mantids engaging in sex with females despite being eaten afterwards. While this behaviour may well be adaptive (Schwartz et al. 2016; Zuk 2016), it is hard to think about such extreme behaviours as involving pain. And yet, larvae - despite their nervous system simplicity - often appear to have richer evaluative capacities than adults, indicative of the different emphasis on survival during this stage, as opposed to reproduction in the adult one. The adult insect body is described by Godfrey-Smith as a mere tool for this end. So it would make sense to have life-stage-dependent varieties of experience. This is something that can straightforwardly be captured within the pathological complexity framework, providing us with different measures at different stages of a life-history.

Barron and Klein (2016) argue for insect consciousness, but largely emphasize the sensory side of things. Yet, findings on the evaluative side are compelling. Godfrey-Smith (2020c) points to the self-administration of analgesics which has been used as compelling evidence of pain in birds and fish, yet has not been found in bees (Groening et al. 2017). But these and other studies have been challenged (Gibbons and Sarlak 2020), leading Browning and Birch (2022) to urge us to be cautious since there is little evidence to expect that morphine must be "a good analgesic in insects" (p. 5). But that is not the only source of evidence we can look for. Bateson et al. (2011) show convincingly that bumblebees, if they have been shaken, can have negative long-term mood states called pessimistic bias. Similarly, Godfrey-Smith (2020c) admits that bees and other insect show aversive responses to heat which may be a better stimulus with which to look for the presence of subjective experience. As I have argued elsewhere, even if insects do not experience human-like pain towards mechanical injuries, they may very well experience other aversive experiences such as hunger or thirst (Browning and Veit 2020b). The absence of pain is too often confused with the absence of evaluation. But the lifestyles of insects simply don't make it necessary to put much of a value on protecting one's bodily shell from mechanical damage. What is important is to complete one's life-history strategy: i.e. to reproduce (or in eusocial insects, to help your sisters with said goal). If wound-tending doesn't aid that, there is little sense in putting much valence on it.

Instead of focusing on pain-like behaviour as an admittedly tempting but flawed paradigm case of evaluative experience, we should look at evidence for a valence-system more generally, one that evaluates conflicting stimuli in a flexible manner. Evidence that is very compelling here, and also highlighted by Godfrey-Smith (2020c), is a follow-up study to Bateson et al. (2011) which focused on positive mood states in the form of optimism bias in response to unexpected sugar rewards in bees (see Perry et al. 2016). The idea of pleasure as a common currency is sometimes criticized as failing to account for the different neural mechanisms of negative and positive evaluation, but such common functional roles of evaluation suggest that they are deeply evolutionary intertwined. Indeed, they *must* largely operate in tandem to allow for efficient decision-making and learning in the face of novel and ambiguous stimuli.

Given how much we know about the level of sensory processing taking place

unconsciously in human brains, it is plausible to think that it is only those sensory inputs that enter the affect system at the final common path of the brain, that are consciously experienced (see Ginsburg and Jablonka 2019 for a similar view). Perhaps a better description would be to see the evolution of 'sensory consciousness' as an enrichment of the evaluative side in respect to its discriminatory capacities. Such a view provides us with an answer to those who insist that functionalist accounts of consciousness cannot explain the 'feel' of experience, since it is precisely this subjective experiencing that does the functional work of dealing with motivational trade-offs; it enables organisms to efficiently deal with their species-specific pathological complexity. So while we could readily admit that insects do not feel pain due to their 'robotic' way of life, their complex behaviours and learning abilities are highly suggestive of something like a common currency of valence for efficient action-selection, even if their evaluative capacities on this side of things may have become poorer compared to their sea-living ancestors. Insects, after all, are in many ways the scaled down versions of their Cambrian ancestors, with a constant pressure for efficiency in organization, especially in those insects that can fly.

Yet, despite (or perhaps precisely because of) this size constraint, the nervous systems of bees have been shown to be incredibly complex. It is because of this that they might provide us with an insight into the minimal nervous system requirements for sentience. Lack of evidence should not here be confused with evidence of absence, precisely because dedicated research on the evaluative capacities (as opposed to their sensory capacities) has been rare. But this is now beginning to change. Indeed, on the 26th of July 2022, a particularly compelling article on motivational trade-offs in bumblebees has been published (see Gibbons et al. 2022). In it, the authors showed that when faced with noxiously-heated feeders and different sugar concentrations, bees were able to trade off competing motivations which they argued could be suggestive of insect pain. Indeed, their work suggests that bumblebees do have a common currency for evaluation. Lastly, recent reviews of insect sentience offer ample additional support for their subjective experience of evaluation (see Mikhalevich and Powell 2020; Lambert et al. 2021).

In conclusion, even what is supposedly the best case for the independent existence of sensory experience without evaluative experience - insects - turns out to demonstrate rich evaluative capacities after all, thus supporting the motivation of the pathological complexity thesis to seek the origins and core of subjective experience in hedonic valence. As Godfrey-Smith (2020b) himself recognizes, it may very well be true that adding a new dimension of consciousness on top of an older more fundamental one may radically transform the phenomenon itself, shifting older features into the background, but that doesn't mean they have disappeared. It is certainly obvious from the human case that the evolution of rich vision can dominate consciousness and we should happily embrace the possibility that subjective experience is as diverse as the diversity of alternative forms of life we find in nature. After all, it is this diversity that ultimately motivates the link between pathological complexity and phenomenological complexity. Let us therefore now think about the other three dimensions and how they may transform consciousness.

5.4 Octopuses: A Disembodied Way of Being

Just as insects challenge our usual thinking about the centrality of pain in evaluative experience, and gastropods challenge our thinking about the need for sensory experience in conscious being, octopuses constitute a challenge to our usual thinking about self-consciousness. What is it like to have a body that can be moved in (figuratively) almost any shape one could want? Their peculiarity and 'alienness' has gained them plenty of attention in discussions of animal minds, and animal welfare legislation in many countries now includes them as 'honorary' vertebrates on the list of species deserving protection. Even the early paper by Eisemann et al. (1984) that denied insect pain, already suggested that "more caution would be needed in other cases, notably that of the cephalopod molluscs, which have a considerably more complex nervous system" (p. 167).[8] Among the molluscs, octopuses stand out for their unique life-history complexity arising from their high degrees of freedom. Furthermore, their soft bodies and foraging lifestyles make them susceptible to predation and parasites, driving up their pathological complexity. Giving up any kind of shell to protect themselves, in addition to possession of an incredible amount of cognitive sophistication and behavioural flexibility, has perhaps made them the most interesting case for an independent evolution of a rich kind of self-consciousness. Like insects, octopuses have short lives, and yet, their bodies appear to be rich and vulnerable investments, rather than the armoured-vehicle-type bodies of insects. Indeed, this combination almost makes them a paradigmatic case for the pathological complexity thesis; vulnerability paired with a high investment requires some sort of conscious control and agency to ensure that this risky use of resources pays off.

Other molluscs might be different in this respect. Walters (2018) notes, for instance, that whereas octopuses have been shown to engage in wound-tending (Alupay et al. 2014), evidential support for the same is lacking in squid, though he also notes that this may simply be an artefact of a lack of studies since "few scientists working in pain-related fields have investigated squid" (p. 7). Nevertheless, there is plenty of convincing evidence that octopuses experience pain-like states, with specialized sensory neurons which appear to focus on noxious stimuli (Alupay et al. 2014; Crook 2021). Here, I am less interested in arguing for the presence of consciousness in these strange but beautiful animals. This case, I think, has been successfully made both on the evaluative and sensory side in recent years. Despite being colour-blind (Messenger 1977; Marshall and Messenger 1996; Bellingham et al. 1998), Birch et al. (2020) describe them as having "rich visual and chemo-tactile perception" (p. 796). Furthermore, following Birch et al. (2020), Mather (2021b) has also written a dedicated article on sensory richness in octopuses, so there is less reason for me to do so here. Likewise, the aforementioned Defra report by Birch et al. (2021) supports a view of very rich evaluative capacities in cephalopod molluscs, including nociception, grooming, and associative learning, among other indicators for evaluative experience. While we could go into much more detail on these two dimensions, this section will instead use octopuses to challenge our preconceived ideas about the dimension of self-consciousness.

[8] Eisemann et al. (1984) cited the work of the influential British octopus researcher Martin Wells (1978) as a potential source for making a case for octopus pain.

Self-Consciousness

Admittedly, I feel some unease in discussing self-consciousness and the unity of experience separately from each other. As I have argued in Chapter 2, these dimensions are incredibly closely related, but since my arguments here do not rely on them being necessarily separate, it won't matter all too much. Octopuses constitute not only an interesting natural experiment for a very different kind of selfhood but also for the possible synchronic disunity of mind. They have often been discussed in this context since their cerebral ganglia and brachial plexus (the nerve ring connecting the neurons in the arms) are arranged in a way that allows some degree of functional autonomy (Godfrey-Smith 2016d, 2019b, 2021a; Carls-Diamante 2017; Mather 2019; Ginsburg and Jablonka 2019). They appear to have a degree of agency in each arm; rather than transferring all information to some sort of central agential control.

While they have both lateralization and centralized integration, their nervous system is best described as *another* strikingly different way of being, perhaps constituting something like a switching of control between arms and the whole organism, rather than two distinct subjects (Godfrey-Smith 2016d). Other admittedly bolder options would be to think about an octopus as having 9 selves - with each individual arm having their own consciousness, their own subjectivity that is only loosely integrated with a higher-level subject: nine selves 'housed' within a single body. Can there be something like a theater of subjects with more or less unified experiences? Do they all have their own perspective or point of view or is this a case where this metaphor breaks down? After all, as Birch et al. (2020) point out: "Octopuses prefer different arms for different activities and their arms appear to function partly independently of the brain" (p. 796). Our folk understanding of consciousness seems to resist such notions, but this may not reveal anything important about the core features of consciousness as a phenomenon in nature.

Since the nervous systems of insects and gastropods may well be sufficient for sentience, we should not be surprised if there could be numerous selves in octopuses, with each of their arms being engaged in their distinct trade-off optimization problems. Within the pathological complexity thesis this would be a sensible prediction to make and we should conduct more empirical work on 'arm-cognition' and information transfer between them to test this hypothesis. We currently lack the evidence to make a confident assertion in either direction, but this is precisely why a comparative approach is needed in order to learn more about the role of *selfhood* and *unity* in consciousness. Importantly, much of the information processing in octopuses is divided between either the right or left parts of their brain (Schnell et al. 2016). But because Mather (2021a) has recently offered a paper discussing synchronic integration in octopuses, it will be doubly useful to focus instead on the experience of a self. Let us thus move through the 'levels' of selfhood and think about their possible evolution by using the lens of the octopus.

A Minimal Sense of a Bodily Self

The evolution of a minimal kind of selfhood, as I argued in Chapter 4, is plausibly found in an enrichment of the sensory side. If consciousness begins with a minimal feeling of evaluation, and later becomes representationally enriched to discriminate more stimuli, it is not at all hard to imagine how further enrichments of just this sort would lead to the feeling of a self in virtue of an eventual distinction between

interoception and exteroception. There does not appear to be a further hard problem here.

What is important to recognize in discussions about the gradual evolution of consciousness and currently existing animals is that many of the gradual steps between the transitions we see today plausibly went extinct. Future evidence may well show that gastropods and insects have a minimal bodily self. Such findings should not be taken as evidence against the pathological complexity thesis. Once an evaluative 'core' of subjective experience appeared in the Cambrian, it is likely that arms races would quickly have led to representational enrichments of this capacity, allowing for better action-control. An awareness of oneself in space would quickly have become useful, similar to the tracking of one's total bodily state. The reason the *state-based* life-histories of these animals are so difficult to model is precisely because different bodily states demand very different actions, and any animal that is able to flexibly alter its behaviour would be able to reap unexplored teleonomic benefits.

Further, I also noted that there is a problem with distinguishing conscious from unconscious self-perception. However, the pathological complexity thesis can here again be used as an answer, by emphasizing only those sensations that enter into the evaluative sphere at a final common path. Most of what goes on in human self-recognition, e.g. balance, walking, and the like, is typically not consciously experienced. It is only under pathological conditions (e.g. damage to the ears), or slippery/moving surfaces (e.g. riding a horse), that our position in space enters our subjective experience, forcing us to consciously balance. But where animals have a very flexible body, we can predict that selfhood will play a more important role in centralized decision-making. Octopuses are able to distinguish their arms from other objects, precisely because their arms could otherwise get in each others' way, which perhaps implies a self-recognition mechanism in each arm to minimize interference (Nesher et al. 2014). Most of this is plausibly unconsciousness, but when conflicts in information from different arms arise there will likely be a need to consciously 'disentangle', and this strongly supports the presence of 'embodied' experience. A minimal degree of this dimension can be expected in all animals with a sufficiently high degree of sensory consciousness.

Self-Awareness

Birch et al. (2020) cite the work by Boal (2006) on social recognition in cephalopods to note that they probably do not possess self-recognition, due to their more or less 'anti-social' lives. Yet, Birch et al. (2020) also cite work on their extraordinary camouflage skills that enable them to hide and pretend to be inanimate objects (Huffard 2006; Hanlon 2007; Hanlon et al. 2011) or imitate noxious animals (i.e. so-called 'fish mimicry'; Hanlon 2007; Hanlon et al. 2010), or clumps of algae (Josef et al. 2012). So the current evidence base may be indicative that the flexible implementation of such camouflage tactics suggests at least some understanding of their own bodies and how others experience them.

Here, it is important to pay very close attention to the life-histories of the animals, to design tests that mimic their typical lifestyles. Given that octopus bodies are so vulnerable and changeable, it is natural to think that their 'body image' is not anything like the stable perception we have of ourselves in space. If a breeze of wind was able to lift up our arms and make them flutter in the wind, we

may be better able to conceive what it is like to have an octopus body, but I do not want to speculate too much here (though see Levy and Hochner 2017). The point is merely to recognize that it is our comparative embodied-ness that is typical for most conscious beings, but that may simply not extend to octopuses in quite the same way (see also Godfrey-Smith 2016d). Where many animals would see a hole and be able to conceive whether they fit through or not, this question is a bit of a trivial matter for an octopus who can fit through a hole much smaller than the size of their body spread out on the ground. What is the sense of having the capacity for self-awareness in the absence of an adaptive rationale? The conflicting evidence is suggestive of perhaps something in-between our usual thinking of self-awareness and minimal embodied phenomenology, but there is no need to think that this would be an experience very similar to our own, or that of other mammals. It might very well be a very alien kind of subjective experience.

Theory of Mind

In thinking about the evolutionary role of the experience of a self, an important function might be the inference from oneself to others, rather than a better organization of one's own sensory inputs. Birch et al. (2020) highlight a recent false-belief task experiment by Kano et al. (2019) that involved orangutans (*Pongo abelii*), chimpanzees (*Pan troglodytes*), and bonobos (*Pan paniscus*) in a test involving both a translucent and an opaque barrier. Interestingly, only those apes that had themselves experienced the barrier as opaque inferred that others would not be able to see objects on the other side of the barrier, thus highlighting the important role of *self-experience* that both human infants and other animals rely on to predict the behaviour of and attribute mental states to others (see also Meltzoff 2007). This research on self-experience is important, because it suggests that "animals can make experience projections, inferences from what they experience in a particular situation to what others will experience" (Birch et al. 2020, p. 797). In addition, there are interesting findings of mindreading abilities in corvids who have been shown to engage in what Birch et al. called *experience projection*, i.e. the prediction of others' behaviour based on their own experience in a similar situation (Emery and Clayton 2001; Dally et al. 2005; Ostojić et al. 2017). Such experiments may be harder to design for octopuses and other sea animals, but that obviously doesn't that mean that they shouldn't be attempted, since they may help us to better understand theory of mind capacities in other animals. Wherever there are complex decisions to be made that involve predicting the knowledge of others it is plausible that these will enter the subjective experience of animals.

Here, it is useful to highlight a recent target article titled "Knowledge before Belief" in *Behavioral and Brain Sciences*, where Phillips et al. (2020) propose a radical reversal of the current paradigm in 'Other Minds' research. Breaking with a long tradition that sought to understand the minds of other humans (and animals) by focusing on the attribution of beliefs, the authors argue that decades of empirical research in the cognitive sciences have undermined, or at least begun to call into question, the assumption that the attribution of knowledge rests on a more basic or fundamental capacity to attribute beliefs. For historical, methodological, and philosophical reasons, however, other minds research has long been held back from even considering this option within the conceptual space.

There is a long-held popular view in philosophy to analyze knowledge as jus-

tified true belief, with belief thus constituting a more basic concept (Ichikawa and Steup 2018). But such a conceptual analysis may not be very convincing when overwhelming evidence from the cognitive sciences appears to tell a different story. Whereas even great apes have consistently been shown in a number of studies to fail at a broad range of false-belief tasks (O Connell and Dunbar 2003; Krachun et al. 2009), Phillips et al. point out that they nevertheless succeeded at representing the knowledge of others (Hare et al. 2000; Melis et al. 2006; Whiten 2013; Karg et al. 2015). If Phillips et al. (2020) are correct, we would be getting a more fine-grained distinction between the representation of others' beliefs (a trait that may be unique to humans) and the representation of others' knowledge (a trait that may be evolutionary ancient). Focusing on human representations of other minds might have biased us once again against a much more basic approach to other minds research.

In a reply to Phillips et al.'s work, I have argued that the ability to track others' knowledge is an evolutionary ancient trait appearing roughly at the beginning of the Cambrian (Veit 2021b). What is notable in the early Cambrian is an increase in body size and the emergence of various sensory modalities to track one's environment. But more sophisticated ways of sensing one's surroundings naturally led to ways of sensing others – to react. This emergence of a richer kind of agency gave rise to arms races between predators and prey (Bengtson 2002) and the evolution of centralized nervous systems (Wray 2015) to better coordinate action and perception.

An important observation made by Godfrey-Smith (2016a) is that there is a transition somewhere in the Cambrian after which "the mind evolved in response to *other* minds" (p. 63). This transition should be understood as the evolution of representing the knowledge within other minds, which can lead to another (though admittedly smaller) explosion in pathological complexity. An important question for both predator and prey becomes: "*Have I been seen?*" The existence of eyes appears to function as a shorthand for many animals to make just this inference – when eyes meet, one infers knowledge of one's location to the subject at the other end of this exchange. Burrowing, camouflage, ink release, and flight are useful attempts to break this link. Many predators avoid the eye contact of their prey at all cost. Knowledge and ignorance of one's surroundings can make all the difference to survival. The evolution of eye-spots on butterflies is one such invention to make potential predators think that they are seen, thus avoiding conflict. Behaviourists may appeal to simpler explanations, but in this case, knowledge attribution may not be such a complex affair.

While octopuses do not engage much with each other, they are active hunters and a desirable source of nutrients to predators. This peculiar life-history - that of a lack of any protective shell, instead involving lots of hiding, camouflage, and tool use (Hanlon et al. 2010; Finn et al. 2009) - is indicative of a special awareness of one's surroundings. The hide-and-seek activities of octopuses at least make sensible the possibility of attributions of an understanding of others' experience. While subjective reports of divers and researchers engaging with octopuses require empirical validation through further experiments, this does not mean that they should not be taken seriously. As Browning (2017) argues, anecdotes can be evidence too and there are plenty of reports of octopuses recognizing different humans, thus suggesting that there is much to learn about these other minds. Some capacities for the recognition of others are likely extremely wide-spread in the animal branch of life, including octopuses, even if these capacities are less sophisticated than in us. Nevertheless, oc-

topuses will present a continuous challenge in our thinking about self-consciousness, with much further work remaining to be done. Their peculiar bodies and lifestyle make them as Schnell and Clayton (2021) argue "suitable ambassadors for rethinking cognition" (p. 27). Let us therefore now move to an exploration of the dimension of synchronic unity, as represented in fishes and non-avian reptiles.

5.5 Fishes and Non-Avian Reptiles: Dis-Unified Ways of Being

In Chapter 3, I mentioned that Birch et al. (2020) suggest using birds as natural experiments of split-brain patients. Yet, they overstate the similarity to split-brain patients since birds have the lower parts of their brain connected, which also makes up a larger fraction of the whole than the cortex does in us (Vallortigara 2000). One would therefore be too quick if one simply treated them as a case for two subjects within a single body - they are more of an intermediate case. Non-avian reptiles - or for that matter, fish - constitute the best neuroanatomical examples for a healthy split-brain because they have (compared to birds) greatly diminished impsilateral projections in the tectofugal and thalamofugal pathways, which make it easier to infer that only one brain 'half' is involved in a task (Deckel 1995, 1997; Vallortigara 2000; Sovrano et al. 2001).

Indeed, the evolution both of birds and mammals[9] is a striking case of an increase in lateral brain connectivity, which is precisely why we should be careful not to overemphasize birds. Their evolution from theropods (the clade of dinosaurs including all predators) is one of consciousness becoming more unified. In thinking about healthy and adaptive forms of disunity we should therefore focus on fish and non-avian reptiles, which I put here together not because they have very similar life-history strategies, but rather because they remain comparatively understudied in comparison to octopuses and birds.

Notably, despite a long history of disregard of the cognitive capacities of fish and non-avian reptiles, due to their status as vertebrates there is now a broad consensus that they are conscious. There are, of course, still those who contest the presence of pain in fish (see Key 2016) and discount the cognitive abilities of non-avian reptiles (see De Meester and Baeckens 2021), but if we want to think about the evolution of synchronic unity within the context of the pathological complexity thesis, these groups constitute the best test cases since we can assume basic evaluative, sensory, and selfhood-related capacities. For excellent reviews of their abilities and capacities linked to consciousness, see Braithwaite (2010); Allen (2013); Woodruff (2017); Sneddon et al. (2018) in the case of fish, and Lambert et al. (2019); Learmonth (2020) in the case of non-avian reptiles. But let us now turn to the capacity for synchronic unity.

Synchronic Experience

If many researchers think that human split-brain patients house two minds, shouldn't we also think of these healthy split-brain organisms as having disunified experience?

[9]Especially in eutherian mammals, who have even more brain connectivity between both halves (see Godfrey-Smith 2021a).

Could it be an efficient way of dealing with the pathological complexity faced by snakes, lizards, and fishes?

As I have argued in Chapter 3, I do not think of the dimensions of unity as genuine ingredients of consciousness. They are rather properties consciousness can have (or for that matter, lack), and I thus see little reason for thinking that they constitute anything like an additional explanatory gap. A certain degree of unity is an automatic feature simply in virtue of informational bottle-necking and the fact that multicellular animals typically act as integrated wholes. A discussion of the evolution of unity (or disunity) must focus on gradual increases or losses in said capacity, not a binary distinction between unity and disunity. Consciousness is highly integrated in us, but that is no reason to think that it must be so in other animals such as fishes and non-avian reptiles, who have very lateralized brains and division of labour between the hemispheres. Does this mean they have two final common paths? Not quite, since performance of most behaviour is by necessity ultimately excluding other behavioural options. Yet, cooperation between brain halves might be a good way to think about lateralized brains. An analogy to tennis players by Godfrey-Smith (2021a) is useful here: just because two agents may work together to form a team, or even something like a collective agency, does not mean that their own perspectives blend into one subjective experience (p. 293). Importantly, not everything that seems to go on in one of the hemispheres is ultimately communicated to the other. There are actual gaps it seems, making it an interesting question to ask why this isn't pathological as it is in the human split-brain case.

Yet, the very question is loaded. To think that lateralized brains are an inferior way of being is a bold assumption, given that the highly lateralized brains of fish and non-avian reptiles are a product of natural selection (i.e. their brains have become more lateralized during their evolution) implying that there must be certain efficiency gains from having a lateralized brain. That human-centric approaches to consciousness may suggest that animals become less conscious through the evolution of lateralization should be counted against those theories, not as evidence for the lack of consciousness. It is a striking fact that vertebrate evolution seemed to have very little selective pressure for *corpus callosum*-like connections between the upper brain hemispheres, which leads others like Godfrey-Smith (2021a) to similarly conclude that such a lateralized way of being can be advantageous. From a neuroeconomics perspective this should admittedly not be all too surprising. There is no way an efficient system can be built (whether biological or artificial) that can make complex decision-making calculations by bringing all information through a single bottleneck. The system would be far too slow, and in the animal case this would mean death. The philosophers' emphasis on agency and goal-directedness routinely leads to the temptation to think of a perfect mind as a Cartesian theater, and we must continuously resist this urge.

Interestingly, Vallortigara and Rogers (2020) follow up this kind of reasoning to argue that the exploitation of asymmetries allows for efficiency gains in vertebrates with lateralized brains. If this is the case, it would explain the preferences of animals for doing particular tasks with one side of their body - not as a mere preference of liking, but as a deep evolutionary efficiency rationale. Drawing on work by Babcock and Robison (1989), Vallortigara and Rogers (2020) argue that lateralization already took place in the Cambrian, when ancestral arthropod-like predators such as *Anomalocaris* "were preying on trilobites with a right-limbed bias" (p. 275). If we were

to add into life-history theory such game-theoretic factors in species-interactions, pathological complexity would dramatically increase - giving us an explanation for why such biases may evolve even in the absence of direct efficiency rationales for the organism. Their adaptive nature might only be understood once we pay close attention to the ecological lifestyles of these animals, which includes knowledge of their predators and prey. Perhaps, a high degree of pathological complexity positively calls for a 'splitting' of consciousness into separate streams.

Godfrey-Smith (2021a) is happy to endorse the idea of two separate streams of cognitive or sensory capacities, but when it comes to the affective side, with evaluative states such as fatigue, he maintains that they are likely shared across the hemispheres - thus giving us partial unity. The apparent difficulty of imagining such a partial split of experience, he attributes to our lack in imaginative capacities; whereas I think these cases pretty easy to think about once we pay close attention to the life-histories of these animals.

Recall our discussion in the preceding chapters of electroencephalograph studies of sleep, that show that animals can rest one brain hemisphere while the other stays active. Does it make sense to think of 'tiredness' as being felt in both sides when the 'restfulness' of each hemisphere can be distinguished? To me it seems that in such cases there is good reason to think that tiredness could exist separately in both brain halves and lead to sleep in one half when it becomes too tired. Imagine a scenario with high predator-load and preferential attention to predators given on one side. Does it not make sense that this side would tire more quickly and that the organism would keep track of that? Other states like hunger may be different since they are better conceived as whole-body states, but the argument for the existence of disunity isn't that every part of experience must be disunified. Animals with disunified brains may be able to flexibly merge their two streams of experience, if they exist, to allow for better coordination, or try to break them apart for better efficiency in particular tasks.

These are inherently speculative suggestions, but the more we learn about split-brains in humans and animals the more realistic these scenarios become. By paying close attention to the life-histories of these animals, I hope, that future work will enable us to better understand the unique adaptive benefits and drawbacks of the integration of experience. As I hope to have demonstrated, simply thinking about the lives of these animals enables us to gradually move away from the human-centric bias in thinking about consciousness. What is pathological in humans, may offer important adaptive benefits to the life-histories of many animals. Much research remains to be done on synchronic unity in these understudied animals, but little sense remains in asserting that brains must be highly integrated, as opposed to lateralized, in order for consciousness to exist. With this in mind, let us turn to the last dimension: diachronic unity.

5.6 Corvids: A Cunning Way of Being

Few animals have been given as much attention in the study of animal cognition as birds. This is unsurprising for two reasons. Firstly, birds are highly intelligent and can often learn experimental procedures quicker than other animals can - thus making them ideal partners in the lab for the study of non-human cognition. Secondly, and perhaps more importantly in the context of this thesis, the observation

of birds was a major cause of both the Darwinian revolution (think of Darwin's finches), and the extension of the Darwinian revolution through the ethologists such as Lorenz (think of his geese), so it is perhaps unsurprising that their study is now also heavily implicated in trying to extend the Darwinian revolution to the mind. As a bird-lover, I give the last substantive discussion in this thesis to them, which I had the pleasure to meet during my time in Nicola Clayton's comparative cognition lab at Cambridge - or to be more precise the group of birds called the *Corvidae* including diverse groups such as crows, jays, magpies, jackdaws, and ravens.

I have already in the discussion of insects emphasized the explosion of pathological complexity with the evolution of flight, and this will be a much more important factor in birds due to their size. Free fall here means death and this may be able to explain the difference in how we treat animal cognition and sentience in birds in comparison to non-avian reptiles. Indeed, corvids appear in many ways akin to us - making them very different cases of non-human consciousness than, say, octopuses. They are long-lived and their life-histories are typically highly social, involving the recognition of the mental states of others (Keefner 2016), similarly also to apes, which makes it hardly surprising that they have a high degree of cognitive complexity. As Schnell and Clayton (2021) acknowledge: "Several current views on the evolution of complex cognition suggest that intelligence coevolved in animals with slow life-histories in response to key socio-ecological challenges" (p. 29). But what makes them particularly interesting in discussions of diachronic unity are their incredible memory capacities, offering us an excellent target system to study the experience of time.

Diachronic Experience

Corvids have been one of the paradigm cases to study mental time travel (i.e. the memory of past experiences and the simulation of future scenarios/experiences), due to their food caching activities. In the longer life-histories of birds, caching makes sense, since food can be stored for future periods such as winter where it might be scarce, thus providing a significant survival benefit against the pathological complexity of bird life, despite requiring an investment into (presumably costly) memory capacities. Clayton and Dickinson (1998) have called the capacity of scrub-jays to remember the what, when, and where, of their caching events 'episodic-like memory' to distinguish it from episodic memory, which is typically defined as conscious memory. Yet, could it be that these birds also have a conscious recollection of their caching activities? Their flexible behaviour may be suggestive of diachronic experience.

Birch et al. (2020) draw on several of studies to suggest that corvids have diachronic experience (pp. 794-795), some of which are worth mentioning here. For instance, Birch et al. mention that Florida scrub jays (*Aphelocoma coerulescens*) have been shown to be able to remember the what, when, and where of their caching activities (Clayton et al. 2001). Furthermore, Birch et al. also note that corvids generally show a great deal of flexibility in the temporal structure of patterns that can be learned, such as that some food will be inedible within a certain time frame, but will be sufficiently 'ripe' after enough time has passed; a form of retrospective cognition and learning (de Kort et al. 2005). Recall from Chapter 2 the paradigm of future planning, where Birch et al. mention the research of Cheke and Clayton (2012) in which Eurasian jays (*Garrulus glandarius*) have been shown to be able

to overcome their current desires in anticipation of future needs and to strategize accordingly, a trait that occurs relatively late in human development. One of the most astonishing findings Birch et al. (2020) mention in this context comes from Kabadayi and Osvath (2017), who demonstrated that ravens (*Corvus corax*) can anticipate how useful tokens or tools can be for future tasks, even when they are entirely novel, thus rivalling the tool-use and bartering capabilities of the great apes.

It thus appears plausible that episodic memory plays an important adaptive role for corvids, enabling them to utilise their intelligence to plan for the future and make use of past memories. What these findings of spontaneous, flexible, and complex planning and anticipation of future needs and desires in response to new information provide, are "promising nonverbal indicator[s] of conscious temporal integration" (Birch et al. 2020, p. 796). For an excellent review of mental time travel in corvids consider Cheke and Clayton (2010). Can these results challenge our human-centric thinking about diachronic unity?

Like Ginsburg and Jablonka (2019), I think that this form of mental time travel constitutes the "evolution of a new type of consciousness that builds on, but goes beyond, the minimal consciousness that we have described up to this point", moving us to what Dennett described as *Popperian creatures* capable of imagination (p. 442). Studies such as these should force us to take seriously the possibility that some corvids might even have richer diachronic experiences than humans, though perhaps not all humans. When it comes to consciousness, there is much more diversity in nature than those working in the science of consciousness would typically like to admit. Those like myself and other philosophers of mind such as Colin Allen[10] suffering from *aphantasia* will surely appreciate this point. We have a reduced, if not absent, ability to engage in the voluntary conjuring of mental images and in sensory-rich mental time travel, while nevertheless having access to semantic memory, i.e. memory of facts, including those about our own past experiences (Zeman et al. 2015; Keogh and Pearson 2018; Pearson 2019). We cannot engage in the reliving or simulation of mental images, smells, touch, and sounds. Yet, we are cognitively capable enough to at least do philosophy. Since these capacities are incredibly complex and play such a central role in the experience of humans one would expect that poorer richness in imagination and mental time travel would lead to losses in some functional capacities. Yet, researchers of the phenomenon were long troubled by the difficulty of finding tests for it, precisely because aphantasics appear to be able to engage in all normal human tasks. The difficulty in detecting aphantasia in human populations may provide good evidence that a high degree of diachronic unity is not necessary for consciousness to be functional.

It should be admitted, however, that research into aphantasia is still in its infancy and we may find better tests in the future to probe the functional capacities of this strange phenomenon, that may well turn out to be pathological. Recent experimental work by Bainbridge et al. (2021) provides compelling evidence for some some impaired functions of aphantasics, through drawing tests relying on memory recall. Aphantasics show "impaired object memory, but intact verbal and spatial memory during recall of real-world scene images. Collectively, these results suggest a dissociation in object and spatial information in visual memory" (Bainbridge et al. 2021, p. 171). Such results are interesting, but there is also some evidence for potential benefits of aphantasia such as "higher memory precision than control participants

[10]Personal communication during my time working as a pre-doc in Pittsburgh.

on some measures, including significantly fewer memory errors and fewer editing in their drawings" (Bainbridge et al. 2021, p. 170). Perhaps this could be related to the memory of facts, which may be easier to correctly remember than experienced events, which are often misreported. Due to a current lack of knowledge about aphantasia, such hypotheses remain speculative, but in the context of evolution it may be tempting to think that improved conceptual and symbolic thinking has enabled humans to become less reliant on conscious memory and thinking. We have truly transcended a stage where we are just Benthamite creatures.

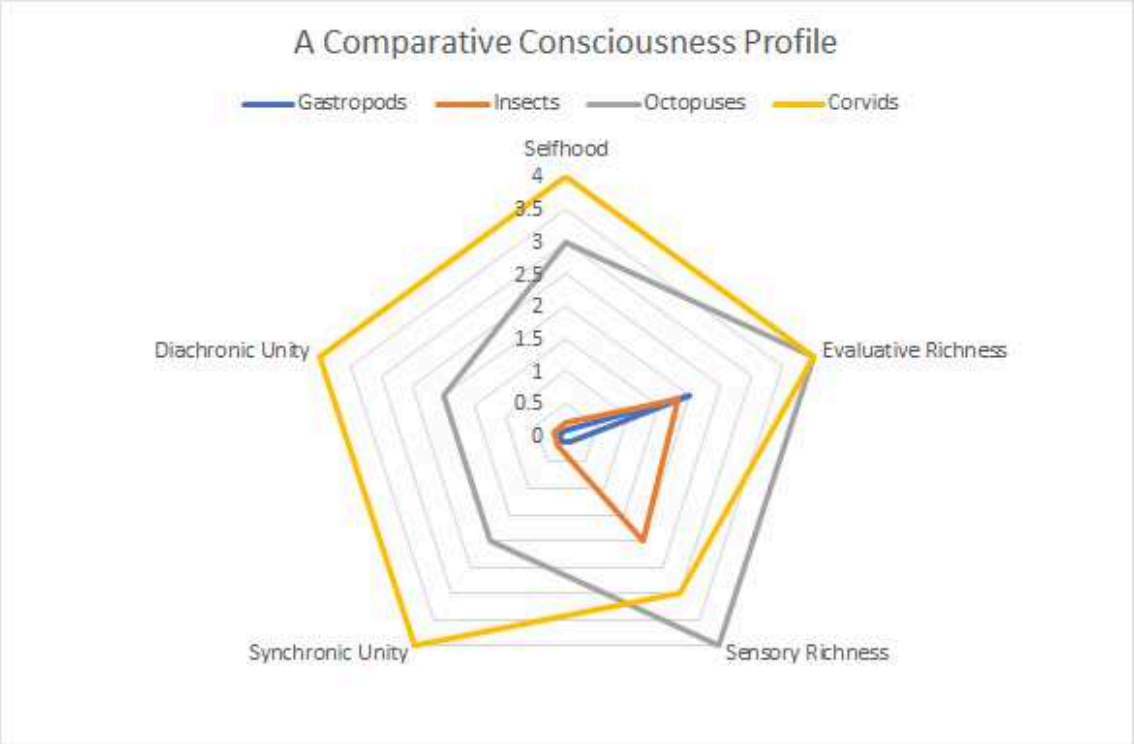
Could corvids similarly rely on semantic memory? Ginsburg and Jablonka (2019) appear to think so, noting that "it may well be that the recall is only richly semantic (the jay has semantic knowledge of where and when and what it cached but does not relive the past experience)" (p. 445). To the contrary, I expect that a certain richness in semantic memory may require higher neural complexity than 'mere' episodic recall. Whereas a human that cannot feel pain will face difficult problems in their lives, they can counteract it in a cognitive manner that would not be available to most animals with that condition - instead they would die. This is why I suspect that consciousness in the sense of felt experiences, rather than rich conceptual thought, is more important to many animals than it is to humans. Instead of thinking about the consciousness profiles depicted in Chapter 2 as having humans covering the richest kinds of each dimension, it may very well be possible that other animals outrank us in more than just smell and touch. And if someone with aphantasia can study mental time travel and visual imagination in other humans, without themselves having this experience, there is no reason to think that we cannot investigate subjective experiences in other animals that may be widely different from what we are familiar with, such as echo-location in bats and infrared sensing in snakes.

Finally, we have to be careful not to be misled by our human conscious experience, an experience that is typically highly integrated and includes episodic memory. Using introspection can tempt one to confuse insights into one's own subjective experience with laws of necessity about experience itself. Perhaps I am 'lucky' here since I fare quite poorly on at least one dimension of the temporal side of things. Philosophy of mind would be much enriched if it recognized that diversity in subjective experience within our species is as much the norm as for any other trait. By mapping out the intra-species variations and gradations of human experience we would make great progress towards recognizing inter-species variations and gradations of consciousness as a phenomenon in nature. The more we learn about this elusive phenomenon and its role in nature, the harder it is to maintain that it is - as Griffin put it - a typologically discrete, all-or-nothing quality unique to a species emerging roughly 300,000 years ago in the African savanna.

5.7 Challenges, Conclusion, and Further Directions

In this penultimate chapter, I had the goal of bringing evolutionary and ecological thinking to the science of consciousness and showing that the other dimensions of consciousness can readily be seen as 'add-ons' - or rather, enrichments - of an evaluative core. This led me to discuss numerous animals such as snails, slugs,

Figure 5.1: Hypothetical consciousness profiles for the phenomenological complexity of gastropods, insects, octopuses, and corvids [modelled after Birch et al. (2020, Figure 1, p. 791)]



fruit flies, crabs, bees, octopuses, fishes, and reptiles of both the avian and non-avian sort. The goal was to demonstrate that we need an ecological and ethological understanding of these animals to develop a truly biological science of consciousness, one that begins with a true understanding of what makes a bat a bat, a snake a snake, and a healthy bee a healthy bee. While this work may appear speculative, we have to seek an understanding of the species-specific pathological complexity of their life-histories, even if many will suspect them to lack sufficient nervous system complexity to warrant an attribution of consciousness.

As I hope to have made clear throughout this thesis, it is precisely with such simple animals that we have to begin an evolutionary investigation of consciousness as a phenomenon that gradually came into existence. The animals usually seen as belonging to a 'grey-zone' are the best cues to what the gradual evolution of subjective experience may have been like, providing us with a rich diversity of different life-history strategies. My discussion of gastropods and arthropods was framed against recent work by Godfrey-Smith (2020b,c), who emphasized the possibility of a disassociation between the sensory and evaluative side in gastropods and insects, which may yield us a two by two table representing capacities across each dimension. Whereas most discussions of consciousness focus on animals that plausibly have together at least a minimal sense of both sensory and evaluative experience, such as most vertebrates and octopuses, a comparative bottom-up approach allows us to reverse-engineer the origins of consciousness by emphasising the animals in which consciousness exists in only a rudimentary form and possibly in a limited subset of dimensions.

Furthermore, by looking at non-human animals we have been able to challenge

assumptions about the other dimensions of consciousness and think about their potential adaptive roles within the life-histories of different animals. In Fig. 5.1 I have offered suggestive comparative consciousness profiles based on the discussions in this chapter. The numbers here are not meant to represent a scale or ordering in a strong sense - only a useful visualization to think about the experiences of these animals. If consciousness can be broken down into different dimensions, we ought to expect a great diversity in the phenomenological complexity of different animals, dependent on their pathological complexity. My goal here was only to walk some of the distance towards a true biological science of consciousness that takes a comparative approach to heart. As Schnell and Clayton (2021) once pointed out, thinking about animal cognition has largely focused on the "challenges experienced by large-brained vertebrates and do not accommodate taxa that have diverse life-histories or that have been exposed to different types of evolutionary pressures" (p. 29). My hope is that the pathological complexity framework can help us to achieve a broadening of scope and that others will join the cause of trying to answer the question of what it is like to be a bee, an octopus, or a raven.

Chapter 6

Steps Towards the Final, Crowning Chapter of the Darwinian Revolution

> *I am confident that with patience and critical investigation we can begin to discern what life is like, subjectively, to particular animals under specific conditions [...]. Cognitive ethologists can certainly improve greatly on these preliminary inferences, once the creative ingenuity of scientists is directed constructively toward the important goal of answering Nagel's basic question: What is it like to be a bat, or any other animal?*
>
> – Donald R. Griffin (1992, p. 260)

6.1 Introduction

We are nearing the end of our journey. Since we have covered a lot of ground thus far, I will conclude this thesis by providing a succinct summary of the core arguments presented, evaluating how far we've come in addressing the goals I set out in the beginning, and offering an optimistic outlook for future investigations.

Chapter Outline

This chapter is structured as follows. In Section 6.2 'Summary', I offer a brief recap of the foregoing chapters. In Section 6.3 'Consciousness Darwinized', I will synthesize these chapters to show how far we have come in making sense of consciousness within a biological view of the mind. Lastly, Section 6.4 'Final Thoughts and Future Directions', will offer tentative suggestions for future research on animal consciousness and the pathological complexity thesis.

6.2 Summary

This thesis opened with an extended introduction to motivate the naturalist project undertaken here. This project aimed to make sense of the place of mind in nature, both in the sense of understanding other minds and the general relationship between mind and matter - two problems to which answers now vary so widely that one may

legitimately think that consciousness is as much, if not more, of a mystery as ever. As I hope to have made clear, however, it is within a *Darwinian standard* that the study of consciousness is provided with a much needed theoretical bottleneck that would constrain and aid the development of a truly biological science of consciousness. This is why, at the beginning of Chapter 1, I placed an epigraph by Griffin in which he maintained that the evolutionary study of consciousness may constitute "the final, crowning chapter of the Darwinian revolution" (1998, p. 14).

Ch. 1: A Darwinian Philosophy for the Science of Consciousness

Following Griffin's call for a cognitive ethology of consciousness, Chapter 1 aimed to make the case that any biological approach to consciousness must address the *teleonomic question* of what consciousness in all of its gradations and varieties does for healthy agents within their normal ecological lifestyles and the natural environments in which they have evolved. To do this, I argued that an evolutionary bottom-up approach to consciousness is both possible and necessary to advance a true Darwinian science of consciousness. But to advance such an approach I also offered my own articulation of a thesis on the function of consciousness - the pathological complexity thesis.

Pathological Complexity Thesis:
The function of consciousness is to enable the agent to respond to pathological complexity.

As the title of my doctoral thesis is meant to emphasize, my motivation for introducing the pathological complexity thesis was to establish a close connection between the natural phenomena of health and consciousness. Firstly, the pathological complexity thesis elaborated the idea that the origin and function of consciousness lies in enabling organisms to efficiently deal with their species-specific pathological complexity, i.e. the trade-offs of their life-histories. To understand health as a natural phenomenon is simply to measure how well an organism succeeds at dealing with their pathological complexity trade-offs, with fitness providing the common currency for biological design. And the origins of consciousness can be understood as the evolution of hedonic valence as a proximate common currency to enable organisms to efficiently calculate these trade-offs in their own actions.

But this was not the only link between health and consciousness that I wanted to emphasize. Drawing on the history of the Darwinian revolution and its extension towards behaviour, I also argued that an extension of this revolution towards consciousness must begin with an appreciation for the distinction between healthy and pathological variations of consciousness. For this, however, we are in need of both a comparative study of consciousness across the animal tree of life and a naturalist understanding of health, which I argued can be derived from our current best theory of organisms. Such a theory will in turn enable us to understand them as active *agents* and *subjects*, including their subjective experience as an integral part of our biological understanding of what makes a bat a bat, a snake a snake, and a healthy bee a healthy bee.

Making use of a number of lessons from the Darwinian revolution, the chapter concluded with an introduction to state-based behavioural and life-history theory as the teleonomic theory of organismal agency and thus the ideal tool with which

to operationalize my concept of pathological complexity. As I hoped to have made clear, it is within this theory that we can bring out the subject-side of organisms and satisfy Lewontin's demand to bring Darwinism to completion, by paying attention to the "functional needs" of the organism (1985, p. 85). Lastly, while important names in the field, such as Godfrey-Smith (2020b) repeatedly speak of a particular animal lifestyle emerging in the Cambrian that led to the evolution of consciousness, this idea has thus far been left rather vague. As a theory of the diversity of 'animal lifestyles', life-history theory is the obvious resource with which to explicate this idea.

Ch. 2: The Explanandum: Animal Consciousness and Phenomenological Complexity

Following my emphasis on the healthy and pathological varieties of consciousness, Chapter 2 argued that any biological theory of consciousness must account for the gradations and varieties of consciousness both within and across species, i.e. what I have dubbed *phenomenological complexity*, as the explanandum of this thesis.

In order to show that a breakdown of consciousness into multiple dimensions, rather than as an all-or-nothing phenomenon, is not only possible but can also be in principle measured in non-human animals, I have reviewed and discussed the call by Birch et al. (2020) to develop multi-dimensional frameworks for animal consciousness, including their own distinction between the five dimensions of sensory experience, unity, temporality, selfhood, and lastly evaluative experience.

By discussing their suggested experimental paradigms to assess these dimensions in animals, in addition to some further ones I introduced, I hope to have removed (or at least weakened) any initial reservations to the possibility of advancing a Darwinian science of consciousness that removes humans from the centre of reference, as was the case with the Darwinian revolution in physiology and behaviour. Both these dimensions and the experimental paradigms associated with them went on to play major roles in the following chapters.

Ch. 3: The Origins of Consciousness or the War of the Five Dimensions

Making use of the distinctions between the five dimensions, Chapter 3 attempted to tease apart the problem of consciousness by treating it like an onion, peeling away one of these dimensions after another, until only one remained: the most plausible contender for the dawn of subjectivity. Rather than treating consciousness as a complex all-or-nothing phenomenon that makes a gradualist evolution seem impossible, I took an evolutionary reverse-engineering approach that tried to narrow the explanatory gap by re-conceiving consciousness as a multi-dimensional phenomenon, in which each dimension could have been built upon one other.

Firstly, I argued that the two dimensions of unity should be seen as ways that experience can be organized, not elements that are necessary for the existence of qualitative experience. The origins of qualia are likely to have been much more akin to fragmented experiences that only later come to be integrated into coherent points of view.

Secondly, I argued that the two dominant traditions in the philosophy of mind and the science of *human* consciousness - strongly externalist representationalist

theories of consciousness that overemphasize sensory experience, and strongly internalist ones that overemphasize self-awareness as the model for all of experience - force us into a false dilemma, with each way of thinking about consciousness seemingly neglecting the other.

While we cannot rule out just yet theories of consciousness built on these dimensions, I argued that the dimension of evaluation can offer us a better dynamic model of consciousness, centred on feedback - one that keeps what is good about both approaches, but discards the rest. Furthermore, it is this dimension that appears to fit better with how laypeople think about consciousness as hedonic valence and to minimize the hard problem challenge, since it is precisely the felt aspects of this capacity that make it functional. And with this rearranged focus on minimal consciousness, the explanatory gap has been significantly reduced.

Unfortunately, Tye (2021) is right to note that "[t]he general topic of the origins of qualia is not one on which philosophers have said a great deal". The goal of my next chapter was thus to address this omission.

Ch. 4: Pathological Complexity and the Dawn of Subjectivity

My goal in Chapter 4 was to offer an explanation for the evolutionary origins of consciousness. It was here that I explicated the pathological complexity thesis in detail.

Firstly, I defended my emphasis on pathological complexity as *the* teleonomic measure of biological complexity that we should focus on - rather than, say, environmental complexity, nervous system complexity, or any other kind of complexity measure we may come up with.

Secondly, I argued that the evolution of hedonic valence can be understood as an adaptive response to a computational explosion of pathological complexity that occurred during the Cambrian, allowing Darwinian agents to track what matters to them and to make the right decisions at the right time. It is in this context that proximate interests or imperative motivations are born, giving rise to a new mode of animal agency in these *Benthamite creatures* with efficient action-control and action-selection.

Lastly, I argued that sensory experience plausibly evolved as an enrichment in the discriminatory capacities of evaluation, whereas minimal selfhood plausibly evolved in further enrichments on the sensory side, involving the distinction between interoception and exteroception. Integration of experience at a time and across time can simply be seen as further transformations in the structure of consciousness, instead of constituting their own explanatory gaps. It is ultimately their being part of an evaluative system that makes these capacities felt.

Ch. 5: Pathological Complexity meets Phenomenological Complexity

In Chapter 5, the pathological complexity thesis was put to the test by responding to an argument by Godfrey-Smith that there could be a phylogenetic split between animals who plausibly only have evaluative experience (i.e. gastropods) and those who only have sensory experience (i.e. insects). Since the pathological complexity thesis rests on the idea that the core and origins of consciousness lie in hedonic evaluation, I used the former group as evidence for my thesis whereas the latter constituted a challenge that had to be overcome.

By reviewing the literature and discussing the life-histories of arthropods, I argued that even those animals that are presently the strongest case for the independent existence of sensory experience turn out to have rich evaluative capacities after all. Furthermore, I discussed octopuses as a challenge to our human-centric thinking about selfhood, fish and non-avian reptiles with their natural split brains as a challenge to synchronic unity, and finally corvids with their spectacular memory capacities as a challenge to our thinking about diachronic experience. In line with the mission of this thesis, these discussions also involved recourse to pathological conditions of consciousness, such as aphantasia.

Finally, the chapter was meant to demonstrate the usefulness of my evolutionary pathological complexity framework by putting the life-histories of animals centre-stage in theorizing about animal consciousness. By linking pathological complexity to phenomenological complexity we are offered a bi-directional approach that can make predictions and feed back into our understanding of the lives of animals, just as the cognitive ethologists originally intended.

6.3 Consciousness Darwinized

While each of the chapters made their own distinctive contributions, and hence could to some degree be read independently, this thesis has largely been one long argument. In order to realize Griffin's vision of a Darwinian science of consciousness, I argued that we must free ourselves from perhaps the most pernicious pre-Darwinian dogma that has held us back for far too long: the faith and confidence in the uniqueness and superiority of the human mind (Griffin 1981, p. 163). For whatever negative impact their tradition still has on the study of animals and consciousness, it was the great insight of the behaviourists of the twentieth century that a psychology that makes consciousness *as we humans experience it* the reference point of all our theories of behaviour, would force us into the same situation that had been faced by biology and medicine prior to the Darwinian revolution. Yet, despite the return of a so-called 'science of consciousness', the field still suffers from just the same flaw as did its precursor in the nineteenth century: a failure to recognize evolutionary continuity. A central motivation of this thesis has been to remove these shackles of a pre-Darwinian view of nature and do for consciousness what the behaviourists have done for behaviour: *to remove humans from our centre of reference*. Instead of accepting this important insight, modern consciousness science operates under the conditions of highly constrained experiments, allowing few degrees of freedom, and studying a only small number of animals (mostly humans).

The pressure to make the science of consciousness as objective as possible, has forced it into a dilemma between externalist and internalist approaches that resembles the pre-Darwinian study of life and behaviour. The problem with a lack of thinking about ecological feedback is that we are missing out on the core of the Darwinian revolution. The absence of ecological thinking in highly constrained experiments is sometimes framed as a conflict between internal validity and ecological validity (i.e. the credibility that our experiment has no bias versus the applicability of our experimental results to the real world), but this obscures what is at stake. In the study of a complex phenomenon, if we frame the problem in this manner then it will always appear to be the rational move to begin with highly constrained situations, but this dictum must be taken with a grain of salt when it comes to

Darwinian phenomena. As I have argued in Chapter 1, the Darwinian revolution did not begin with the study of animals in the lab - rather it came out of natural history, i.e. the observation of healthy and pathological cases of living activity in a natural setting. This is why the ethologists maintained that an extension of the Darwinian revolution to behaviour must begin here: with an understanding of the natural healthy lives of animals.

We need to ground our scientific theorizing firmly within an ecological understanding of the organism. To demand experimental paradigms occurring prior to this stage to fit within the straightjacket of highly controlled experiments, such as those conducted in physics, would have severely held back the modern biology of behaviour. But whereas the research areas of animal behaviour and cognition now have a healthy pluralism of accepted methods and experiments - combining fieldwork, ecological models, and constrained experiments in the lab - the study of consciousness still suffers from the old conflict between 'comparative' psychology and the truly comparative methods of ethology. The demand for scientific 'objectivity' has put consciousness science into this straightjacket of highly controlled experiments, something that is doubly problematic if, as I have argued here, consciousness evolved to deal with the pathological complexity caused by high degrees of freedom. This is the important lesson we should take away from the ethologists' Darwinian revolution in studying behaviour: in order to study a biological phenomenon we must use a truly comparative approach that attends to the similarities and differences in octopuses, bees, bats, insects, and the like by investigating what it means to be a healthy agent of that species acting in the normal ecological setting in which it evolved. And such a Darwinian approach must also recognize a multiplicity of gradations and variations in the subjective experience both within and across different species, which I have called phenomenological complexity.

To begin the final, crowning chapter of the Darwinian revolution must mean to learn from the Darwinian revolution of life and to force the problems of consciousness through this theoretical bottleneck of evolutionary theory. This is the standard that the study of consciousness was so desperately lacking. As I have hoped to show in this thesis, modern evolutionary biology - and in particular state-based life-history theory - offers us rich theoretical resources and constraints within which to think about the problems of mind. At the forefront of a biological science of consciousness must be the functionalist question of what consciousness in all of its diversity and gradations does *for* healthy sentient agents within their normal ecological lifestyles and the natural environments in which they have evolved. And in order to answer this question, we must attend to the very origins of this phenomenon, rather than its special presentation in humans. A truly Darwinian comparative approach cannot make the human case the model for all consciousness.

Such a science of consciousness will inevitably transform our understanding of the phenomenon, similarly to how the science of space and time have done so for their target phenomena.[1] Philosophers such as Nagel are skeptical of this idea, since they think that our understanding of consciousness must be anchored in our first-person experience. But a true biological science of consciousness needs only to honour our experience as one variety of consciousness; a special case of a much more widespread and diverse phenomenon. Just as theories of space-time should account for how space and time appear for humans as a special case of a more

[1] I thank Paul Griffiths for this suggested analogy.

general phenomenon, so can a science of consciousness transform our understanding of this phenomenon as one that is complex and multi-dimensional, with plenty of varieties and gradations of phenomenological complexity. As Bryce Huebner and I have argued in a commentary in *Animal Sentience*, the "future of animal sentience research lies not in drawing boundaries but in empirically investigating *what it feels like to be* an echo-locating bat, an infrared-sensing snake, an octopus with multiple distributed ganglia, a fish without a neocortex, or an arthropod such as a spider or a honey bee" (Veit and Huebner 2020, p. 3).

Not only have I argued that we must begin with a teleonomic understanding of the organism, i.e. what organisms are *for*, which will enable us to make distinctions between the normal and the pathological, but I have also argued that consciousness itself is an adaptive response to complex health challenges arising at a multicellular level from an active animal lifestyle, that constitutes a significant evolutionary transition in agency. Health, I argued, can be understood as a measure of how well an organism deals with its pathological complexity, which I have argued in turn can be explicated in terms of modern state-based behavioural and life-history theory. Pathological complexity corresponds to the decision-theoretic fitness maximization problem of life-history agents trying to secure their presentation in the next generation. Fitness provides us with a common currency for comparing the health status of different animals, whether they have lesions, parasites, broken bones, diseases, or poor environments. Health and pathology, as Lorenz argued early on, must be understood in an ecological context. And it is in the context of the Cambrian explosion and the evolution of behavioural flexibility - of predation, fluctuating environments, and more complex lifestyles - that pathological complexity exploded, making the proximate common currency of valence a requirement in order for organisms to efficiently map their behaviour to environmental states, depending on their own bodily state. The origins of consciousness, as speculated by Darwin's protege Romanes and many other evolutionists, lies in the dawn of affective feelings, of the evaluation of states and behaviours as good, bad, or neutral.

While the available evidence cannot rule out competing theories of consciousness just yet, my theory has two major points in its favour, giving it an edge over many other contenders in the field. Firstly, the 'hard problem' of why things feel a certain way does not appear to be *as much* of a challenge within a hedonic framework, as opposed to a strongly externalist representationalist one. Things feel a certain way, because they *have to* feel that way to be functional, which also provides a superior picture to strongly internalist views that make consciousness an almost automatic feature. A focus on evaluation is able to overcome the alleged inability of functionalist theories of consciousness to explain why consciousness *feels* like anything at all (see also Solms 2021). Secondly, it is within such a model of evaluation as the basis of consciousness that the other dimensions of consciousness can readily be explained as later transformations. As I have shown in Chapter 5, the pathological complexity thesis allows us to make predictions regarding the subjective experiences of other animals such as insects, octopuses, and corvids - predictions that can be tested with the experimental paradigms discussed in Chapter 2. Phenomenological complexity can be explained as an adaptive response to pathological complexity. This is a second major advantage of my framework against most competitors that have a hard time making testable predictions, such as the Integrated Information Theory of consciousness. And the empirical implications of the pathological complexity thesis

appear to at least provide us with something like a good initial test for its viability. Further empirical tests will be able to push it further, allowing for the integration of subjects and *subjectivity* into the Darwinian revolution, thus moving us towards Griffin's goal of creating a cognitive ethology that makes sense of the place of mind in nature through recourse to a sound evolutionary understanding of an animal's evolved life-history strategies.

6.4 Final Thoughts and Future Directions

Because of skepticism against the very idea of animal minds, researchers were long forced to make their experiments as empirically rigorous and controlled as possible, at the cost of telling us little about what the mind does *for* animals in the wild. As I noted above, this omission can be attributed to a very old and still entrenched difference between comparative psychologists and cognitive ethologists that maps onto an even older conflict between behaviourists and the classical ethologists. While both groups can be interested in the evolution of traits, those in the tradition of comparative cognition are faced with a "constant pressure towards laboratory experimentation in conditions that are often of questionable ecological validity" (Allen 2004, p. 605) and largely focus on a small range of species. Cognitive ethologists, on the other hand, emphasize the importance of observing a broad range of taxonomic groups in conditions as close as possible to the ecological conditions in which they have evolved (see also Allen 2006).

Naturally, such an approach makes experiments hard, if not impossible, in many cases. So there are inevitable trade-offs here. Nevertheless, ecological field experiments and observations can be useful in thinking about the possible adaptive value of consciousness in animals. But this could hardly be further from anything like a true test for the presence of consciousness, a demand that I hope we will be able to overcome in the near future. While the dividing lines between these traditions have weakened, they are still very much alive when it comes to the study of animal consciousness, and the two will have to become integrated for a true biological science of consciousness to emerge. As Allen and Bekoff (1997) have argued early on, we are in need of *species-fair* tests that are adapted to the different lifestyles of different taxa (p. xiv). Tests for diverse species must include sound evolutionary and ecological understanding of their lives in order to avoid premature conclusions about what they can or cannot experience, and I hope that the pathological complexity thesis can help us in thinking about the design of such tests.

The evolutionary account I have tried to sketch here necessarily had gaps. In doing philosophy of nature, one can not be concerned with the precise nor the causal specifics of any particular situation. Instead, one draws on and synthesizes evidence from a variety of different disciplines to make sense of the world on a very general level - in our case the place of mind in nature, both in the sense of the connection between matter and mind, and the literal locations of conscious agents in the world. In trying to develop a general framework for this task - a model for thinking about consciousness in a bottom-up fashion across the tree of life - there are inevitable theoretical trade-offs. But that is just what has to be provided at the beginnings of a true biological science of a complex phenomenon. The question is not how a particular bat is conscious, but how consciousness itself can be made sense of in the most general form as a phenomenon in nature. This general approach is not

anti-neural as much as it is functionalist. We are asking for the teleonomic reason for the existence of subjective experience.

Future work will inevitably improve upon the framework I have tried to provide here. But none of this will be a problem if it can serve as a theoretical scaffold for future bottom-up approaches that take Darwinian thinking seriously, emphasizing both gradations and varieties of subjective experience in animal life. What has been shown, I hope, is that this framework is not just a 'just-so story'. All living agents have to deal with the complex design-problem of trading off different functional demands against each other. Just as fitness constitutes an ultimate common currency for evaluating the health status of an organism, so does hedonic valence provide a proximate common currency for agents to deal with a degree of pathological complexity that would otherwise become unmanageable for a behaviourally flexible, vulnerable, multicellular organism. Consciousness evolved to enable animals to efficiently deal with this pathological complexity by providing a dynamic way of evaluating trade-offs in the deployment of alternative actions.

One curious feature of this account is that this explosion of pathological complexity may in principle have occurred in separate lineages, so just as there are several independent origins of behavioural complexity, there could be multiple independent origins of subjective experience. Godfrey-Smith (2020b) does not attempt to offer a conclusive answer but raises the possibility that consciousness may have independently arisen in cephalopods, arthropods, and vertebrates. Some locate consciousness only in vertebrates or mammals, because of their similarities to us, but this view is making less sense the more we learn about other branches of life. The lifestyles of cephalopods and crustaceans are too similar to those of many vertebrates to deny the adaptive value of sentience in these creatures, and there is now plenty of experimental evidence that supports the attribution of pain to them (Birch et al. 2021). Whether this common evaluative core arose once or multiple times is an open question, but the model presented here can make sense of an early occurrence of this capacity as a plausible candidate scenario. As a much more basic capacity required to even allow for such behavioural flexibility to become manageable for the organism, it would explain the difference between the Avalon and Cambrian explosions.

However, while I have offered an alternative to views associating consciousness with later evolutionary innovations, such as rich forms of information integration or self-hood, this doesn't mean proponents of such views must be entirely wrong. Indeed, I see it as a virtue of my account that their claims regarding a more recent dawn of consciousness may very well be reconceived of instead as what Godfrey-Smith (2016d,c) has described as a *transformation view* as opposed to a *latecomer view*. A transformation view accepts that "some late-evolving features of our brains do greatly *affect* the nature of subjective experience", but does not treat them as bringing consciousness into existence (Godfrey-Smith 2016c, p. 499). The transformation view alongside the pathological complexity thesis makes the problem of minimal consciousness significantly smaller and narrows the explanatory gap through the use of the hedonic picture, but it also has the virtues of accommodating the evidence presented by latecomer views. While latecomer views can give the impression of being less speculative by asking for a great degree of similarity between other animals and us in order to grant them consciousness, they commit the fallacy of failing to distinguish human consciousness from consciousness as a phenomenon in

nature, which plausibly has a much simpler origin.

Indeed, if we want to hold onto both the idea that consciousness has only arisen once and that it is shared by these three lineages, it must have arisen in a common ancestor that was very simple. And as far as I have argued here, the pathological complexity thesis makes more sense of such an early occurrence than any of its competitors (such as those accounts provided by Godfrey-Smith or Ginsburg and Jablonka) that emphasize later features arising independently in these three lineages. Somewhere around the beginning of the Cambrian explosion, if not earlier, lived the last common ancestor of arthropods and vertebrates. It is plausibly here, in some worm-like creature resembling modern annelids, that we find something like the very first sparks of feelings. However, I do not want to assert that all descendants of these semi-proto-quasi sentient creatures must be conscious. As I noted in my discussion of insects, some branches of animal life could have become more simple and 'lost' sentience. After all, as I have argued throughout this dissertation, consciousness must ultimately pay off for organisms, and simpler lives with lower pathological complexity may well succeed without hedonic valence.

The pathological complexity thesis was ultimately intended as a teleonomic theory of consciousness, but it may also offer us some clues regarding the metaphysics of consciousness. In particular, we should plausibly think of consciousness in terms of something like a hybrid materialist-functionalist identity thesis that identifies the origin of consciousness with the origin of valence systems in multicellular animals, somewhere around the beginning of the Cambrian. To be such an animal simply means to be a subject with experience. Tracing the gradual evolution of Benthamite creatures from objects into subjects/agents makes the mystery of consciousness substantially smaller that it would otherwise seem.

Perhaps the most important consequence of the ideas in this thesis is the relevance of the connections between pathological complexity and sentience for the purposes of ethics and policy-making, since it is sentience that is usually taken to make an entity a subject of moral concern (Browning 2020c; Browning and Veit 2022c). Some dub this view 'sentientism' - the view that "you have moral status, i.e. you are a subject of moral concern, if and only if you are sentient, i.e. if and only if you are capable of phenomenally consciously experiencing pleasure or pain" (Sebo 2018, p. 4). As my interest in animal consciousness was also motivated by ethical concerns, I have published several papers on animal ethics, sentience, and welfare science in collaboration with Heather Browning,[2] but one problem we were repeatedly faced with is the challenge of inter-species comparisons of welfare: that is, the more broadly we attribute sentience to other animals, the less reasonable it will be to assign equal moral weight to all insects, birds, and octopuses (see Browning 2022b for a detailed examination). Their capacity to suffer and experience pleasure reasonably scales according to their degree of consciousness, which forces us to think about consciousness in a gradualist, rather than 'on' or 'off' manner. While we have offered some brief discussions on animal sentience in relation to the life-histories of different animals (Browning and Veit 2021c), the pathological complexity thesis may offer us a useful evolutionary proxy measure to assess different levels of evaluative richness in the subjective experience of different animals. How to measure animal welfare is a notoriously difficult problem (Browning 2022a), but a measure of pathological complexity would enable us (at least in principle) to assess animals

[2]See Veit and Browning (2020a, 2021); Browning and Veit (2020c,a, 2021b, 2022a).

according to a so-called *sentience multiplier* in calculations for how to best improve animal welfare (see Browning 2020a), and this could in turn be tested with the experimental paradigms for evaluative richness discussed in Chapter 2. But this will be a task for future work.

A similar path may be opened for discussions of consciousness in robots, the presence of which I believe would require conditions and abilities approximating the pathological complexity of animal lifestyles - something that has not yet been achieved. It may well be that the particular material basis of evaluation in animal agents matters for consciousness, but the functionalist approach offered here could at least offer us clues for how to build sentient machines. To understand consciousness we will first have to understand its evolutionary origins, rather than speculate about consciousness in non-biological entities. Other future applications of the framework may also extend to the explanation of disorders and varieties of consciousness in humans. Since my being on both the autism and aphantasia spectrums has informed my awareness of the phenomenological diversity of minds, I hope to expand into this area in the future.[3]

Finally, I hope there is some truth in the account I have offered and that my methodology of reverse-engineering the origins of consciousness within hedonic evaluation and the pathological complexity of animal life has moved us a few steps closer towards the writing of the final, crowning chapter of the Darwinian revolution.

[3]I already have a few publications on the topic of autism (Chapman and Veit 2021, 2020; Browning and Veit forthcoming), but I hope in the future to apply my framework to the evolution of mind-reading abilities.

Bibliography

Adam, R. and O. Güntürkün (2009). When one hemisphere takes control: meta-control in pigeons (Columba livia). *PLoS One 4*(4), e5307.

Adamo, S. A. (2016). Do insects feel pain? A question at the intersection of animal behaviour, philosophy and robotics. *Animal Behaviour 118*, 75–79.

Albrecht, G. L. and P. J. Devlieger (1999). The disability paradox: high quality of life against all odds. *Social Science & Medicine 48*(8), 977–988.

Allen, C. (2004). Is anyone a cognitive ethologist? *Biology & Philosophy 19*(4), 589–607.

Allen, C. (2006). Transitive inference in animals: Reasoning or conditioned associations. In S. Hurley and M. Nudds (Eds.), *Rational Animals?*, pp. 175–185. Oxford: Oxford University Press.

Allen, C. (2013). Fish cognition and consciousness. *Journal of Agricultural and Environmental Ethics 26*(1), 25–39.

Allen, C. (2017). Associative learning. In K. Andrews and J. Beck (Eds.), *The Routledge Handbook of Animals Minds*, pp. 401–408. New York: Routledge.

Allen, C. and M. Bekoff (1997). *Species of Mind: The Philosophy and Biology of Cognitive Ethology*. Cambridge, MA: MIT Press.

Allen, C., P. N. Fuchs, A. Shriver, and H. D. Wilson (2005). Deciphering Animal Pain. In M. Aydede (Ed.), *Pain: New Essays on Its Nature and the Methodology of Its Study*, pp. 351–366. Cambridge, MA: MIT Press.

Allen, C. and M. Trestman (2017). Animal Consciousness. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 ed.). https://plato.stanford.edu/archives/win2017/entries/consciousness-animal/ [Accessed on 22/05/2022].

Alupay, J. S., S. P. Hadjisolomou, and R. J. Crook (2014). Arm injury produces long-term behavioral and neural hypersensitivity in octopus. *Neuroscience Letters 558*, 137–142.

Anderson, J. R. (1986). Mirror-mediated finding of hidden food by monkeys (Macaca tonkeana and M. fascicularis). *Journal of Comparative Psychology 100*(3), 237–242.

Anderson, J. R. and G. G. Gallup (2015). Mirror self-recognition: a review and critique of attempts to promote and engineer self-recognition in primates. *Primates 56*(4), 317–326.

Animal Ethics (2020). Establishing a research field in natural sciences: three case studies. *Oakland: Animal Ethics*, 1–62. https://www.animal-ethics.org/establishing-field-naturalsciences [Accessed on 25/03/2021].

Antolin, M. F. (2011). Evolution, medicine, and the Darwin family. *Evolution: Education and Outreach 4*(4), 613–623.

Appel, M. and R. W. Elwood (2009). Motivational trade-offs and potential pain experience in hermit crabs. *Applied Animal Behaviour Science 119*(1-2), 120–124.

Archer, J. (1992). *Ethology and Human Development*. Herfordshire: Harvester Wheatsheaf.

Aristotle (1991). *On the Soul (J. Barnes, Ed., J. A. Smith, Trans.). The Complete Works of Aristotle: The Revised Oxford Translation, vol. 1.* Princeton, NJ: Princeton University Press.

Aveling, E. B. (1886). *Die Darwin'sche Theorie*. Stuttgart: Verlag von J. H. W. Dietz.

Avramides, A. (2020). Other Minds. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020 ed.). https://plato.stanford.edu/archives/win2020/entries/other-minds/ [Accessed on 17/04/2022].

Baars, B. (1988). *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Babcock, L. E. and R. A. Robison (1989). Preferences of Palaeozoic predators. *Nature 337*(6209), 695–696.

Bainbridge, W. A., Z. Pounder, A. F. Eardley, and C. I. Baker (2021). Quantifying Aphantasia through drawing: Those without visual imagery show deficits in object but not spatial memory. *Cortex 135*, 159–172.

Balasko, M. and M. Cabanac (1998a). Behavior of juvenile lizards (Iguana iguana) in a conflict between temperature regulation and palatable food. *Brain, Behavior and Evolution 52*(6), 257–262.

Balasko, M. and M. Cabanac (1998b). Motivational conflict among water need, palatability, and cold discomfort in rats. *Physiology & Behavior 65*(1), 35–41.

Balduzzi, D. and G. Tononi (2008). Integrated information in discrete dynamical systems: motivation and theoretical framework. *PLoS Computational Biology 4*(6), e1000091.

Balduzzi, D. and G. Tononi (2009). Qualia: the geometry of integrated information. *PLoS Computational Biology 5*(8), e1000462.

Balleine, B. W. and A. Dickinson (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology 37*(4-5), 407–419.

Bard, K. A., B. K. Todd, C. Bernier, J. Love, and D. A. Leavens (2006). Self-awareness in human and chimpanzee infants: What is measured and what is meant by the mark and mirror test? *Infancy 9*(2), 191–219.

Barlassina, L. (2020). Beyond good and bad: Reflexive imperativism, not evaluativism, explains valence. *Thought: A Journal of Philosophy 9*(4), 274–284.

Barlassina, L. and M. K. Hayward (2019). More of me! Less of me!: Reflexive imperativism about affective phenomenal character. *Mind 128*(512), 1013–1044.

Baron-Cohen, S. (1997). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.

Baron-Cohen, S. (2000). Theory of mind and autism: A review. *International Review of Research in Mental Retardation 23*, 169–184.

Barrett, L. F. (2017). *How Emotions Are Made: The Secret Life of the Brain*. London: Pan Macmillan.

Barron, A. B. and C. Klein (2016). What insects can tell us about the origins of consciousness. *Proceedings of the National Academy of Sciences 113*(18), 4900–4908.

Barron, A. B., E. Søvik, and J. L. Cornish (2010). The roles of dopamine and related compounds in reward-seeking behavior across animal phyla. *Frontiers in Behavioral Neuroscience 4*, 163.

Bateson, M., S. Desire, S. E. Gartside, and G. A. Wright (2011). Agitated honeybees exhibit pessimistic cognitive biases. *Current Biology 21*(12), 1070–1073.

Beer, C. (2020). Niko Tinbergen and questions of instinct. *Animal Behaviour 164*, 261–265.

Bellingham, J., A. G. Morris, and D. M. Hunt (1998). The rhodopsin gene of the cuttlefish Sepia officinalis: sequence and spectral tuning. *Journal of Experimental Biology 201*(15), 2299–2306.

Bengtson, S. (2002). Origins and early evolution of predation. *The Paleontological Society Papers 8*, 289–318.

Bentham, J. (1879). *An Introduction to the Principles of Morals and Legislation*. London: T. Payne & Son.

Benton, M. J. and P. C. Donoghue (2007). Paleontological evidence to date the tree of life. *Molecular Biology and Evolution 24*(1), 26–53.

Bergson, H. (1920). *Mind-Energy: Lectures and Essays*. New York: Henry Holt. (W. Carr, Trans.).

Bermond, B. (2001). A neuropsychological and evolutionary approach to animal consciousness and animal suffering. *Animal Welfare 10*(1), 47–62.

Berridge, K. C. (1996). Food reward: brain substrates of wanting and liking. *Neuroscience & Biobehavioral Reviews 20*(1), 1–25.

Berridge, K. C. (2009a). Wanting and liking: Observations from the neuroscience and psychology laboratory. *Inquiry 52*(4), 378–398.

Berridge, K. C. (2009b). 'Liking' and 'wanting' food rewards: Brain substrates and roles in eating disorders. *Physiology & Behavior 97*(5), 537–550.

Billard, P., N. S. Clayton, and C. Jozet-Alves (2020). Cuttlefish retrieve whether they smelt or saw a previously encountered item. *Scientific Reports 10*(1), 1–7.

Birch, J. (2020a). In search of the origins of consciousness. *Acta Biotheoretica 68*(68), 287–294.

Birch, J. (2020b). The search for invertebrate consciousness. *Noûs*.

Birch, J., D. M. Broom, H. Browning, A. Crump, S. Ginsburg, M. Halina, D. Harrison, E. Jablonka, A. Y. Lee, F. Kammerer, et al. (2022). How should We study animal consciousness scientifically? *Journal of Consciousness Studies 29*(3-4), 8–28.

Birch, J., C. Burn, A. Schnell, H. Browning, and A. Crump (2021). *Review of the Evidence of Sentience in Cephalopod Molluscs and Decapod Crustaceans*. London: LSE Consulting.

Birch, J., S. Ginsburg, and E. Jablonka (2020). Unlimited Associative Learning and the origins of consciousness: a primer and some predictions. *Biology & philosophy 35*(6), 1–23.

Birch, J., A. K. Schnell, and N. S. Clayton (2020). Dimensions of animal consciousness. *Trends in Cognitive Sciences 24*(10), 789–801.

Blackstone, N. W. (2001). Crustacea (Crustaceans). *eLS*. https://doi.org/10.1002/9780470015902.a0001606.pub3.

Block, N. (2004). Qualia. In R. L. Gregory (Ed.), *Oxford Companion to the Mind*, pp. 785–789. Oxford: Oxford University Press.

Boal, J. (2006). Social recognition: a top down view of cephalopod behaviour. *Vie et Milieu 56*(2), 69–80.

Bonner, J. T. (1988). *The Evolution of Complexity*. Princeton, NJ: Princeton University Press.

Boorse, C. (1977). Health as a theoretical concept. *Philosophy of science 44*(4), 542–573.

Bowler, P. (1992). *The Non-Darwinian Revolution: Reinterpreting a Historical Myth*. Johns Hopkins University Press.

Bowler, P. J. (2013). *Darwin deleted: imagining a world without Darwin*. University of Chicago Press.

Braithwaite, V. (2010). *Do Fish Feel Pain?* Oxford: Oxford University Press.

Brandon, R. (1990). *Adaptation and Environment*. Princeton, NJ: Princeton University Press.

Bronfman, Z. Z., S. Ginsburg, and E. Jablonka (2016). The transition to minimal consciousness through the evolution of associative learning. *Frontiers in Psychology 7*, 1954.

Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence 47*, 139–159.

Broom, D. M. (2014). *Sentience and Animal Welfare*. Wallingford: CABI.

Browning, H. (2017). Anecdotes can be evidence too. *Animal Sentience 16*(13).

Browning, H. (2019). What should we do about sheep? The role of intelligence in welfare considerations. *Animal Sentience 25*(23).

Browning, H. (2020a). Assessing measures of animal welfare. *Preprint*. http://philsci-archive.pitt.edu/17144/.

Browning, H. (2020b). *If I Could Talk to the Animals: Measuring Subjective Animal Welfare*. Ph.D. thesis, Australian National University.

Browning, H. (2020c). The natural behavior debate: Two conceptions of animal welfare. *Journal of Applied Animal Welfare Science*, 325–337.

Browning, H. (2022a). The measurability of subjective animal welfare. *Journal of Consciousness Studies 29*(3-4), 150–179.

Browning, H. (2022b). The problem of interspecies welfare comparisons. *Preprint*. http://philsci-archive.pitt.edu/20115/.

Browning, H. and J. Birch (2022). Animal sentience. *Philosophy Compass*, e12822.

Browning, H. and W. Veit (2020a). Confined freedom and free confinement: The ethics of captivity in *Life of Pi*. In Á. T. Bogár and R. S. Szigethy (Eds.), *Critical Insights: Life of Pi*, pp. 119–134. Amenia, NY: Salem Press.

Browning, H. and W. Veit (2020b). Improving invertebrate welfare. *Animal Sentience 29*(4).

Browning, H. and W. Veit (2020c). Is humane slaughter possible? *Animals 10*(5), 799.

Browning, H. and W. Veit (2020d). The measurement problem of consciousness. *Philosophical Topics 48*(1), 85–108.

Browning, H. and W. Veit (2021a). Evolutionary biology meets consciousness: essay review of Simona Ginsburg and Eva Jablonka's *The Evolution of the Sensitive Soul*. *Biology & Philosophy 36*(5). https://doi.org/10.1007/s10539-021-09781-7.

Browning, H. and W. Veit (2021b). Freedom and animal welfare. *Animals 11*(4), 1148.

Browning, H. and W. Veit (2021c). Positive wild animal welfare. *Preprint*. http://philsci-archive.pitt.edu/19608/.

Browning, H. and W. Veit (2022a). The importance of end-of-life welfare. *Animal Frontiers 12*(1), 8–15.

Browning, H. and W. Veit (2022b). On the relevance of experimental philosophy to neuroethics. *AJOB Neuroscience 13*(1), 55–57.

Browning, H. and W. Veit (2022c). The sentience shift in animal research. *The New Bioethics*, 1–16. https://doi.org/10.1080/20502877.2022.2077681.

Browning, H. and W. Veit (forthcoming). Autism and the preference for imaginary worlds. *Behavioral and Brain Sciences*.

Brunberg, E., A. Wallenbeck, and L. J. Keeling (2011). Tail biting in fattening pigs: Associations between frequency of tail biting and other abnormal behaviours. *Applied Animal Behaviour Science 133*(1-2), 18–25.

Bublitz, A., S. R. Weinhold, S. Strobel, G. Dehnhardt, and F. D. Hanke (2017). Reconsideration of serial visual reversal learning in octopus (Octopus vulgaris) from a methodological perspective. *Frontiers in Physiology 8*, 54.

Budd, G. E. and M. J. Telford (2009). The origin and evolution of arthropods. *Nature 457*(7231), 812–817.

Burgdorf, J. and J. Panksepp (2006). The neurobiology of positive emotions. *Neuroscience & Biobehavioral Reviews 30*(2), 173–187.

Burghardt, G. M. (1985). Animal awareness: Current perceptions and historical perspective. *American Psychologist 40*(8), 905–919.

Cabanac, M. (1971). Physiological role of pleasure. *Science 173*(4002), 1103–1107.

Cabanac, M. (1979). Sensory pleasure. *The Quarterly Review of Biology 54*(1), 1–29.

Cabanac, M. (1992). Pleasure: the common currency. *Journal of Theoretical Biology 155*(2), 173–200.

Cabanac, M. (1996). On the origin of consciousness, a postulate and its corollary. *Neuroscience & Biobehavioral Reviews 20*(1), 33–40.

Cabanac, M. (1999). Emotion and phylogeny. *Journal of Consciousness Studies 6*(6-7), 176–190.

Cabanac, M., A. J. Cabanac, and A. Parent (2009). The emergence of consciousness in phylogeny. *Behavioural Brain Research 198*(2), 267–272.

Cabanac, M. and K. Johnson (1983). Analysis of a conflict between palatability and cold exposure in rats. *Physiology & Behavior 31*(2), 249–253.

Cacioppo, J. T. and W. L. Gardner (1999). Emotion. *Annual Review of Psychology 50*(1), 191–214.

Calcott, B. (2009). Lineage explanations: explaining how biological mechanisms change. *The British Journal for the Philosophy of Science*.

Call, J. and M. Tomasello (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences 12*(5), 187–192.

Calvo, P., M. Gagliano, G. M. Souza, and A. Trewavas (2020). Plants are intelligent, here's how. *Annals of Botany 125*(1), 11–28.

Campbell, N. (2001). What was Huxley's epiphenomenalism? *Biology & Philosophy 16*(3), 357–375.

Canguilhem, G. (1991). *The Normal and the Pathological*. New York: Zone Books. Trans. C. R. Fawcett.

Caporael, L. R., J. R. Griesemer, and W. C. Wimsatt (2014). *Developing Scaffolds in Evolution, Culture, and Cognition*. Cambridge, MA: MIT Press.

Carel, H. (2007). Can I be ill and happy? *Philosophia 35*(2), 95–110.

Carew, T. J., E. T. Walters, and E. R. Kandel (1981). Associative learning in Aplysia: Cellular correlates supporting a conditioned fear hypothesis. *Science 211*(4481), 501–504.

Carls-Diamante, S. (2017). The octopus and the unity of consciousness. *Biology & Philosophy 32*(6), 1269–1287.

Carnap, R. (1950). *Logical Foundations of Probability*. Chicago, IL: University of Chicago Press.

Carruthers, P. (2018). Valence and value. *Philosophy and Phenomenological Research 97*(3), 658–680.

Carruthers, P. (forthcoming). On Valence: Imperative or Representation of Value? *The British Journal for the Philosophy of Science*.

Caveney, S., W. Cladman, L. Verellen, and C. Donly (2006). Ancestry of neuronal monoamine transporters in the Metazoa. *Journal of Experimental Biology 209*(24), 4858–4868.

Cazzolla Gatti, R., A. Velichevskaya, B. Gottesman, and K. Davis (2021). Grey wolf may show signs of self-awareness with the sniff test of self-recognition. *Ethology Ecology & Evolution 33*(4), 444–467.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies 2*(3), 200–219.

Chalmers, D. J. (2020). Is the hard problem of consciousness universal? *Journal of Consciousness Studies 27*(5-6), 227–257.

Chapman, R. and W. Veit (2020). Representing the autism spectrum. *The American Journal of Bioethics 20*(4), 46–48.

Chapman, R. and W. Veit (2021). "The essence of autism: fact or artefact?". *Molecular Psychology 26*(5), 1440–1441.

Cheke, L. G. and N. S. Clayton (2010). Mental time travel in animals. *Wiley Interdisciplinary Reviews: Cognitive Science 1*(6), 915–930.

Cheke, L. G. and N. S. Clayton (2012). Eurasian jays (Garrulus glandarius) overcome their current desires to anticipate two distinct future needs and plan for them appropriately. *Biology Letters 8*(2), 171–175.

Christensen, M. S., L. Kristiansen, J. B. Rowe, and J. B. Nielsen (2008). Action-blindsight in healthy subjects after transcranial magnetic stimulation. *Proceedings of the National Academy of Sciences 105*(4), 1353–1357.

Churchland, P. S. (1988). Reduction and the neurobiological basis of consciousness. In A. J. Marcel and E. Bisiach (Eds.), *Consciousness in Contemporary Science*, pp. 273–304. Oxford: Oxford University Press.

Churchland, P. S. (2002). *Brain-Wise: Studies in Neurophilosophy*. Cambridge, MA: Cambridge, MA: MIT Press.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Clark, R. E., J. R. Manns, and L. R. Squire (2002). Classical conditioning, awareness, and brain systems. *Trends in Cognitive Sciences 6*(12), 524–531.

Clayton, N. S. and A. Dickinson (1998). Episodic-like memory during cache recovery by scrub jays. *Nature 395*(6699), 272–274.

Clayton, N. S., K. S. Yu, and A. Dickinson (2001). Scrub jays (Aphelocoma coerulescens) form integrated memories of the multiple features of caching episodes. *Journal of Experimental Psychology: Animal Behavior Processes 27*(1), 17.

Colwill, R. M., R. A. Absher, and M. Roberts (1988). Context-US learning in Aplysia californica. *Journal of Neuroscience 8*(12), 4434–4439.

Cowey, A. (2010). The blindsight saga. *Experimental Brain Research 200*(1), 3–24.

Cox, J. J., F. Reimann, A. K. Nicholas, G. Thornton, E. Roberts, K. Springell, G. Karbani, H. Jafri, J. Mannan, Y. Raashid, et al. (2006). An SCN9A channelopathy causes congenital inability to experience pain. *Nature 444*(7121), 894–898.

Crick, F. and C. Koch (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences 2*, 263–275.

Crook, R. J. (2021). Behavioral and neurophysiological evidence suggests affective pain experience in octopus. *iScience 24*(3), 102229.

Crook, R. J. and E. T. Walters (2011). Nociceptive behavior and physiology of molluscs: animal welfare implications. *ILAR Journal 52*(2), 185–195.

Crump, A., G. Arnott, and E. J. Bethell (2018). Affect-driven attention biases as animal welfare indicators: review and methods. *Animals 8*(8), 136.

Dale, R. and J. M. Plotnik (2017). Elephants know when their bodies are obstacles to success in a novel transfer task. *Scientific Reports 7*, 46309.

Dally, J. M., N. J. Emery, and N. S. Clayton (2005). Cache protection strategies by western scrub-jays, Aphelocoma californica: implications for social cognition. *Animal Behaviour 70*(6), 1251–1263.

Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Avon Books.

Damasio, A. R. (1999). *The Feeling Of What Happens: Body, Emotion and the Making of Consciousness*. New York: Harcourt Brace.

Danckert, J., C. Striemer, and Y. Rossetti (2021). Blindsight. In J. J. S. Barton and A. Leff (Eds.), *Handbook of Clinical Neurology*, Volume 178, pp. 297–310. Elsevier.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. London: John Murray.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London: John Murray.

Darwin, C. and A. Wallace (1858). On the Tendency of Species to form Varieties and on the Perpetuation of Varieties and Species by Natural Means of Selection. *Journal of the Proceedings of the Linnean Society of London. Zoology 3*(9), 45–62.

Darwin, C. R. (1838). Notebook M: [Metaphysics on morals and speculations on expression]. *Darwin Online*. CUL-DAR125. Transcribed by Kees Rookmaaker, edited by Paul Barrett.

Darwin, C. R. (1840). Old & useless notes about the moral sense & some metaphysical points. *Darwin Online*. CUL-DAR91.4-55. Transcribed and edited by Paul H. Barrett.

Darwin, E. (1794). *Zoonomia; or, the Laws of Organic Life (Volume 1)*. London: Johnson.

Dawkins, M. (2017a). Animal welfare with and without consciousness. *Talk at NYU 2017 Animal Consciousness Conference*. `https://www.youtube.com/playlist?list=PLY_s7b9LrR8UCcPqL59XuIII68ALjgLt7` [Accessed on 12/03/2021].

Dawkins, M. S. (1990). From an animal's point of view: motivation, fitness, and animal welfare. *Behavioural and Brain Sciences 13*(1), 1–9.

Dawkins, M. S. (1998). Evolution and animal welfare. *The Quarterly Review of Biology 73*(3), 305–328.

Dawkins, M. S. (2001). Who needs consciousness? *Animal Welfare 10*, S19–29.

Dawkins, M. S. (2017b). Animal welfare with and without consciousness. *Journal of Zoology 301*(1), 1–10.

Dawkins, M. S. (2021). *The Science of Animal Welfare: Understanding What Animals Want*. Oxford: Oxford University Press.

de Kort, S. R., A. Dickinson, and N. S. Clayton (2005). Retrospective cognition by food-caching western scrub-jays. *Learning and Motivation 36*(2), 159–176.

De Meester, G. and S. Baeckens (2021). Reinstating reptiles: from clueless creatures to esteemed models of cognitive biology. *Behaviour 158*(12-13), 1057–1076.

De Waal, F. (2016). *Are We Smart Enough to Know How Smart Animals Are?* New York: WW Norton & Company.

de Waal, F. B. (2019). Fish, mirrors, and a gradualist perspective on self-awareness. *PLoS Biology 17*(2), e3000112.

de Waal, F. B. and P. F. Ferrari (2010). Towards a bottom-up perspective on animal and human cognition. *Trends in Cognitive Sciences 14*(5), 201–207.

D'Eath, R. B. (1998). Can video images imitate real stimuli in animal behaviour experiments? *Biological Reviews 73*(3), 267–292.

Deckel, A. W. (1995). Laterality of aggressive responses in *Anolis*. *Journal of Experimental Zoology 272*(3), 194–200.

Deckel, A. W. (1997). Effects of alcohol consumption on lateralized aggression in *Anolis carolinensis*. *Brain Research 756*(1-2), 96–105.

Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. New York: Viking.

Dennett, D. (1973). Mechanism and responsibility. In T. Honderich (Ed.), *Essays on Freedom of Action*, pp. 159–184. London: Routledge.

Dennett, D. (1991). *Consciousness Explained*. New York: Little, Brown and Co.

Dennett, D. (2019a). Review of Other Minds: the octopus, the sea and the deep origins of consciousness. *Biology & Philosophy 34*(1).

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: MIT press.

Dennett, D. C. (1995a). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon and Schuster.

Dennett, D. C. (1995b). The selfish gene as a philosophical essay. In A. Grafen and M. Ridley (Eds.), *Richard Dawkins: How a Scientist Changed the Way We Think*, pp. 101–115. Oxford: Oxford University Press.

Dennett, D. C. (1996). *Kinds of Minds: Toward an Understanding of Consciousness*. Cambridge, MA: MIT Press.

Dennett, D. C. (2005). *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: MIT Press.

Dennett, D. C. (2017a). *From Bacteria to Bach and Back: The Evolution of Minds*. New York: WW Norton & Company.

Dennett, D. C. (2017b). Suffering matters. *Talk at NYU 2017 Animal Consciousness Conference*. `https://www.youtube.com/playlist?list=PLY_s7b9LrR8UCcPqL59XuIII68ALjgLt7` [Accessed on 12/03/2021].

Dennett, D. C. (2019b). Welcome to strong illusionism. *Journal of Consciousness Studies 26*(9-10), 48–58.

Denton, D. (2006). *The Primordial Emotions: The Dawning of Consciousness.* Oxford: Oxford University Press.

Diamond, J. M. (1982). Big-bang reproduction and ageing in male marsupial mice. *Nature 298*(5870), 115–116.

Dobson, F. S. (2013). Live fast, die young, and win the sperm competition. *Proceedings of the National Academy of Sciences 110*(44), 17610–17611.

Dörrenberg, S., H. Rakoczy, and U. Liszkowski (2018). How (not) to measure infant theory of mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development 46*, 12–30.

Dretske, F. (1993). Conscious experience. *Mind 102*(406), 263–283.

Dretske, F. (1999). The mind's awareness of itself. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 95*(1/2), 103–124.

Dunlap, K. (1919). Are there any instincts? *Journal of Abnormal Psychology 14*, 307–311.

Edelman, G. M. and G. Tononi (2000). *A Universe of Consciousness: How Matter Becomes Imagination*. New York: Basic Books.

Edwards, A. (2007). Maximisation principles in evolutionary biology. In M. Matthen and C. Stephens (Eds.), *Handbook of the Philosophy of Science: Philosophy of Biology*, pp. 335–347. Amsterdam: North-Holland.

Eisemann, C., W. Jorgensen, D. Merritt, M. Rice, B. Cribb, P. Webb, and M. Zalucki (1984). Do insects feel pain?—A biological view. *Experientia 40*(2), 164–167.

Elwood, R. W. et al. (2012). Evidence for pain in decapod crustaceans. *Animal Welfare 21*(S2), 23–27.

Elwood, R. W. and M. Appel (2009). Pain experience in hermit crabs? *Animal Behaviour 77*(5), 1243–1246.

Emery, N. J. and N. S. Clayton (2001). Effects of experience and social context on prospective caching strategies by scrub jays. *Nature 414*(6862), 443–446.

Ereshefsky, M. (2009). Defining 'health'and 'disease'. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 40*(3), 221–227.

Escobar, J. M. (2012). Autopoiesis and Darwinism. *Synthese 185*(1), 53–72.

Feinberg, T. and J. Mallatt (2016). *The Ancient Origins of Consciousness*. Cambridge, MA: MIT press.

Figdor, C. (2018). *Pieces of Mind: The Proper Domain of Psychological Predicates*. Oxford: Oxford University Press.

Figdor, C. (2020). Relationship between cognition and moral status needs overhaul. *Animal Sentience 29*(3).

Finn, J. K., T. Tregenza, and M. D. Norman (2009). Defensive tool use in a coconut-carrying octopus. *Current Biology 19*(23), R1069–R1070.

Flanagan, O. J. (1991). *The Science of the Mind*. Cambridge, MA: MIT Press.

Frankish, K. (2017). *Illusionism: as a Theory of Consciousness*. Exeter: Imprint Academic.

Fritz, R. S., N. E. Stamp, and T. G. Halverson (1982). Iteroparity and semelparity in insects. *The American Naturalist 120*(2), 264–268.

Gagliano, M. (2017). The mind of plants: thinking the unthinkable. *Communicative & Integrative Biology 10*(2), 38427.

Gagliano, M. (2018). Inside the vegetal mind: on the cognitive abilities of plants. In *Memory and Learning in Plants*, pp. 215–220. Cham: Springer.

Gagliano, M., V. V. Vyazovskiy, A. A. Borbély, M. Grimonprez, and M. Depczynski (2016). Learning by association in plants. *Scientific Reports 6*(1), 1–9.

Gallagher, S. (2011). *The Oxford Handbook of the Self*. New York: Oxford University Press.

Gallup, G. G. (1970). Chimpanzees: Self-recognition. *Science 167*(3914), 86–87.

Gallup Jr, G. G. and J. R. Anderson (2020). Self-recognition in animals: Where do we stand 50 years later? Lessons from cleaner wrasse and other species. *Psychology of Consciousness: Theory, Research, and Practice 7*(1), 46.

Gardner, A. (2009). Adaptation as organism design. *Biology Letters 5*(6), 861–864.

Gardner, A. (2017). The purpose of adaptation. *Interface Focus 7*(5), 20170005.

Gardner, A. (2019). The agent concept is a scientific tool. *Metascience 28*(3), 359–363.

Gibbons, M. and S. Sarlak (2020). Inhibition of pain or response to injury in invertebrates and vertebrates. *Animal Sentience 5*(29), 34.

Gibbons, M., E. Versace, A. Crump, B. Baran, and L. Chittka (2022). Motivational trade-offs and modulation of nociception in bumblebees. *Proceedings of the National Academy of Sciences 119*(31), e2205821119.

Ginsburg, S. and E. Jablonka (2010). Experiencing: a Jamesian approach. *Journal of Consciousness Studies 17*(5-6), 102–124.

Ginsburg, S. and E. Jablonka (2019). *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness*. Cambridge, MA: MIT Press.

Godfrey-Smith, P. (1996a). *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.

Godfrey-Smith, P. (1996b). Précis of Complexity and the Function of Mind in Nature. *Adaptive Behavior 4*(3-4), 453–465.

Godfrey-Smith, P. (1996c). Replies to four critics. *Adaptive Behavior 4*(3-4), 486–493.

Godfrey-Smith, P. (1997). Author's response. *Metascience 6*(2), 31–37.

Godfrey-Smith, P. (2002). Environmental complexity and the evolution of cognition. In R. Sternberg and J. Kaufman (Eds.), *The Evolution of Intelligence*, pp. 223–249. Mahwah, NJ: Lawrence Erlbaum.

Godfrey-Smith, P. (2013a). On the relation between philosophy and science. *Unpublished lecture manuscript from the Gesellschaft für Wissenschaftsphilosophie (GWP) Conference, Hannover 2013.*. `https://www.petergodfreysmith.com/PhilosophyScience_PGS_2013.pdf` [Accessed on 23/06/2020].

Godfrey-Smith, P. (2013b). *Philosophy of Biology*. Princeton, NJ: Princeton University Press.

Godfrey-Smith, P. (2016a). Animal evolution and the origins of experience. In D. Livingstone Smith (Ed.), *How Biology Shapes Philosophy: New Foundations for Naturalism*, pp. 51–71. Cambridge: Cambridge University Press.

Godfrey-Smith, P. (2016b). Individuality, subjectivity, and minimal cognition. *Biology & Philosophy 31*(6), 775–796.

Godfrey-Smith, P. (2016c). Mind, matter, and metabolism. *The Journal of Philosophy 113*(10), 481–506.

Godfrey-Smith, P. (2016d). *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness*. New York: Farrar, Straus and Giroux.

Godfrey-Smith, P. (2017a). Animal evolution and subjective experience. *Talk at NYU 2017 Animal Consciousness Conference*. `https://www.youtube.com/playlist?list=PLY_s7b9LrR8UCcPqL59XuIII68ALjgLt7` [Accessed on 12/03/2021].

Godfrey-Smith, P. (2017b). Complexity revisited. *Biology & Philosophy 32*(3), 467–479.

Godfrey-Smith, P. (2017c). The subject as cause and effect of evolution. *Interface Focus 7*(5), 20170022.

Godfrey-Smith, P. (2019a). Evolving across the explanatory gap. *Philosophy, Theory, and Practice in Biology 11*.

Godfrey-Smith, P. (2019b). Octopus experience. *Animal Sentience 26*(18).

Godfrey-Smith, P. (2020a). Gradualism and the evolution of experience. *Philosophical Topics 48*(1), 201–220.

Godfrey-Smith, P. (2020b). *Metazoa: Animal Life and the Birth of the Mind*. New York: Farrar, Strauß, and Giroux.

Godfrey-Smith, P. (2020c). Varieties of Subjectivity. *Philosophy of Science 87*(5), 1150–1159.

Godfrey-Smith, P. (2021a). Integration, lateralization, and animal experience. *Mind & Language 36*(2), 285–296.

Godfrey-Smith, P. (2021b). Learning and the biology of consciousness: a commentary on Birch, Ginsburg, and Jablonka. *Biology & Philosophy 36*(5), 1–4.

Goff, P., W. Seager, and S. Allen-Hermanson (2020). Panpsychism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020 ed.). https://plato.stanford.edu/archives/sum2020/entries/panpsychism/ [Accessed on 17/02/2021].

Goldman, A. I. (1993). Consciousness, folk psychology, and cognitive science. *Consciousness and Cognition 2*(4), 364–382.

Gould, S. J. (1996). In the Mind of the Beholder. In *Dinosaur in a Haystack*, pp. 93–107. London: Jonathan Cape.

Gould, S. J. and R. C. Lewontin (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of The Royal Society of London. Series B. Biological Sciences 205*(1161), 581–598.

Grafen, A. (2009). Formalizing Darwinism and inclusive fitness theory. *Philosophical Transactions of the Royal Society B: Biological Sciences 364*(1533), 3135–3141.

Grafen, A. (2014). The formal darwinism project in outline. *Biology & Philosophy 29*(2), 155–174.

Graziano, M. S. (2016). Consciousness engineered. *Journal of Consciousness Studies 23*(11-12), 98–115.

Griesemer, J. (2014). Reproduction and scaffolded developmental processes: an integrated evolutionary perspective. In A. Minelli and T. Pradeu (Eds.), *Towards a Theory of Development*, pp. 183–202. Oxford: University Press Oxford.

Griffin, D. R. (1976). *The Question of Animal Awareness: Evolutionary Continuity of Mental Experience*. New York: Rockefeller University Press.

Griffin, D. R. (1981). *The Question of Animal Awareness: Evolutionary Continuity of Mental Experience (Revised and Enlarged Edition)*. New York: Rockefeller University Press.

Griffin, D. R. (1992). *Animal Minds*. Chicago, IL: University of Chicago Press.

Griffin, D. R. (1998). From cognition to consciousness. *Animal Cognition 1*(1), 3–16.

Griffin, D. R. (2001). *Animal Minds: Beyond Cognition to Consciousness*. Chicago, IL: University of Chicago Press.

Griffiths, P. E. (1997). *What Emotions Really Are: The Problem of Psychological Categories*. Chicago, IL: University of Chicago Press.

Griffiths, P. E. (2008). Ethology, sociobiology, and evolutionary psychology. In S. Sarkar and A. Plutynski (Eds.), *A Companion to the Philosophy of Biology*, pp. 393–414. Hoboken, NJ: Blackwell Publishing Ltd.

Griffiths, P. E. (2009). In what sense does 'nothing make sense except in the light of evolution'? *Acta Biotheoretica 57*(1), 11–32.

Griffiths, P. E. (2017). Emotions. In W. Bechtel and G. Graham (Eds.), *A Companion to Cognitive Science*, pp. 197–203. Hoboken, NJ: Blackwell Publishing Ltd.

Griffiths, P. E. (2018). What is an organism, and what is it for? Neo-Aristotelian, Darwinian and post-Hamilton perspectives. *Talk at PhilInBioMed Network Bordeaux*. `https://youtu.be/TPuYjYT_eKo` [Accessed on 20/02/2021].

Griffiths, P. E. and R. D. Gray (2001). Darwinism and developmental systems. In S. Oyama, P. Griffiths, and R. D. Gray (Eds.), *Cycles of Contingency: Developmental Systems and Evolution*, pp. 195–218. Cambridge, MA: MIT Press.

Griffiths, P. E. and J. Matthewson (2018). Evolution, dysfunction, and disease: A reappraisal. *The British Journal for the Philosophy of Science 69*(2), 301–327.

Groening, J., D. Venini, and M. V. Srinivasan (2017). In search of evidence for the experience of pain in honeybees: A self-administration study. *Scientific Reports 7*(1), 1–8.

Güntürkün, O. and T. Bugnyar (2016). Cognition without cortex. *Trends in Cognitive Sciences 20*(4), 291–303.

Haeckel, E. (1892). Our monism. The principles of a consistent, unitary world-view. *The Monist 2*(4), 481–486.

Hanlon, R. (2007). Cephalopod dynamic camouflage. *Current Biology 17*(11), R400–R404.

Hanlon, R. T., C. Chiao, L. Mäthger, K. C. Buresch, A. Barbosa, J. J. Allen, L. Siemann, and C. Chubb (2011). Rapid adaptive camouflage in cephalopods. In M. Stevens and S. Merilaita (Eds.), *Animal camouflage: Mechanisms and Functions*, pp. 145–163. Cambridge: Cambridge University Press.

Hanlon, R. T., A. C. Watson, and A. Barbosa (2010). A "mimic octopus" in the Atlantic: flatfish mimicry and camouflage by Macrotritopus defilippi. *The Biological Bulletin 218*(1), 15–24.

Hare, B., J. Call, B. Agnetta, and M. Tomasello (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour 59*(4), 771–785.

Hayden, B. Y. and Y. Niv (2021). The case against economic values in the orbitofrontal cortex (or anywhere else in the brain). *Behavioral Neuroscience 135*(2), 192.

Healy, K., L. McNally, G. D. Ruxton, N. Cooper, and A. L. Jackson (2013). Metabolic rate and body size are linked with perception of temporal information. *Animal behaviour 86*(4), 685–696.

Hentschel, E., , and H. Penzlin (1982). Beeinflussung des Putzverhaltens bei *Periplaneta americana* (L.) durch Wundsetzung, Naloxon-, Morphin-und Met-Enkephalingaben. *Zoologische Jahrbücher. Abteilung fur allgemeine Zoologie und Physiologie der Tiere 86*, 361–370.

Hesslow, G. (1993). Do we need a concept of disease? *Theoretical Medicine 14*(1), 1–14.

Hill, C. S. (2018). Unity of consciousness. *Wiley Interdisciplinary Reviews: Cognitive Science 9*(5), e1465.

Ho, M.-W. and P. T. Saunders (1979). Beyond Neo-Darwinism—an epigenetic approach to evolution. *Journal of theoretical Biology 78*(4), 573–591.

Houston, A. I. and J. M. McNamara (1999). *Models of Adaptive Behaviour: An Approach Based on State*. Cambridge: Cambridge University Press.

Huebner, B. (2012). Reflection, reflex, and folk intuitions. *Consciousness and Cognition 21*(2), 651–653.

Huffard, C. L. (2006). Locomotion by Abdopus aculeatus (Cephalopoda: Octopodidae): walking the line between primary and secondary defenses. *Journal of Experimental Biology 209*(19), 3697–3707.

Humphrey, N. (1992). *A History of the Mind: Evolution and the Birth of Consciousness*. New York: Simon and Schuster.

Humphrey, N. (2011). *Soul Dust: The Magic of Consciousness*. Princeton, NJ: Princeton University Press.

Hurley, S. L. (1998). *Consciousness in Action*. Cambridge, MA: Harvard University Press.

Huxley, T. (2011). On the Hypothesis that Animals Are Automata, and Its History. In *Collected Essays. Volume 1: Methods and Results*, pp. 199–250. Cambridge: Cambridge University Press.

Huxley, T. H. and W. J. Youmans (1868). *The Elements of Physiology and Hygiene: A Text-book for Educational Institutions*. New York: Appleton & Co.

Ichikawa, J. J. and M. Steup (2018). The Analysis of Knowledge. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018 ed.). https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/ [Accessed on 18/01/2020].

James, W. (1890). *The Principles of Psychology*. New York: Dover Publications.

Jékely, G., F. Keijzer, and P. Godfrey-Smith (2015). An option space for early neural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences 370*(1684), 20150181.

Jennings, H. S. (1904). *Behavior of the Lower Organisms*. New York: Columbia University Press.

Jolij, J. and V. A. Lamme (2005). Repression of unconscious information by conscious processing: evidence from affective blindsight induced by transcranial magnetic stimulation. *Proceedings of the National Academy of Sciences 102*(30), 10747–10751.

Josef, N., P. Amodio, G. Fiorito, and N. Shashar (2012). Camouflaging in a complex environment—octopuses use specific features of their surroundings for background matching. *PLoS One 7*(5), e37579.

Kabadayi, C. and M. Osvath (2017). Ravens parallel great apes in flexible planning for tool-use and bartering. *Science 357*(6347), 202–204.

Kano, F., C. Krupenye, S. Hirata, M. Tomonaga, and J. Call (2019). Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proceedings of the National Academy of Sciences 116*(42), 20904–20909.

Kappeler, P. M. (2021). *Animal Behaviour: An Evolutionary Perspective*. Cham: Springer.

Karg, K., M. Schmelz, J. Call, and M. Tomasello (2015). The goggles experiment: can chimpanzees use self-experience to infer what a competitor can see? *Animal Behaviour 105*, 211–221.

Keefner, A. (2016). Corvids infer the mental states of conspecifics. *Biology & Philosophy 31*(2), 267–281.

Keeley, B. L. (2002). Making sense of the senses: Individuating modalities in humans and other animals. *The Journal of Philosophy 99*(1), 5–28.

Keenan, J. P., G. G. Gallup, and D. Falk (2003). *The Face in the Mirror: The Search for the Origins of Consciousness*. New York: HarperCollins Publishers.

Keijzer, F. (2015). Moving and sensing without input and output: early nervous systems and the origins of the animal sensorimotor organization. *Biology & Philosophy 30*(3), 311–331.

Keijzer, F. and A. Arnellos (2017). The animal sensorimotor organization: a challenge for the environmental complexity thesis. *Biology & philosophy 32*(3), 421–441.

Keijzer, F., M. Van Duijn, and P. Lyon (2013). What nervous systems do: early evolution, input–output, and the skin brain thesis. *Adaptive Behavior 21*(2), 67–85.

Keogh, R. and J. Pearson (2018). The blind mind: No sensory visual imagery in aphantasia. *Cortex 105*, 53–60.

Key, B. (2016). Why fish do not feel pain. *Animal Sentience 1*(3), 1–34.

Kirk, R. (2021). Zombies. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021 ed.). https://plato.stanford.edu/archives/spr2021/entries/zombies/ [Accessed on 04/12/2021].

Klein, C. and A. B. Barron (2020). How experimental neuroscientists can fix the hard problem of consciousness. *Neuroscience of Consciousness 2020*(1), niaa009.

Knobe, J. and S. Nichols (2017). Experimental philosophy. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 ed.). https://plato.stanford.edu/archives/win2017/entries/experimental-philosophy/ [Accessed on 28/02/2021].

Koch, C. and G. Tononi (2011). A test for consciousness. *Scientific American 304*(6), 44–47.

Kohda, M., T. Hotta, T. Takeyama, S. Awata, H. Tanaka, J. Asai, and A. L. Jordan (2019). If a fish can pass the mark test, what are the implications for consciousness and self-awareness testing in animals? *PLoS Biology 17*(2), e3000021.

Kohda, M., S. Sogawa, A. L. Jordan, N. Kubo, S. Awata, S. Satoh, T. Kobayashi, A. Fujita, and R. Bshary (2022). Further evidence for the capacity of mirror self-recognition in cleaner fish and the significance of ecologically relevant marks. *PLoS Biology 20*(2), e3001529.

Krachun, C., M. Carpenter, J. Call, and M. Tomasello (2009, July). A competitive nonverbal false belief task for children and apes. *Developmental Science 12*(4), 521–535.

Kringelbach, M. and K. Berridge (2010). *Pleasures of the Brain*. Oxford: Oxford University Press.

Krupenye, C. and J. Call (2019). Theory of mind in animals: Current and future directions. *Wiley Interdisciplinary Reviews: Cognitive Science 10*(6), e1503.

Kuo, Z. Y. (1921). Giving up instincts in psychology. *Journal of Philosophy 18*, 645–664.

Kutschera, U., G. S. Levit, and U. Hossfeld (2019). Ernst Haeckel (1834–1919): The German Darwin and his impact on modern biology. *Theory in Biosciences 138*(1), 1–7.

Labandeira, C. C. (2006). Silurian to Triassic plant and insect clades and their associations: new data, a review, and interpretations. *Arthropod Systematics & Phylogeny 64*, 53–94.

Lack, D. (1947). The significance of clutch-size. *Ibis 89*(2), 302–352.

Laland, K., T. Uller, M. Feldman, K. Sterelny, G. B. Müller, A. Moczek, E. Jablonka, J. Odling-Smee, G. A. Wray, H. E. Hoekstra, et al. (2014). Does evolutionary theory need a rethink? *Nature News 514*(7521), 161.

Lamarck, J. (1984). *Zoological Philosophy*. Chicago: University of Chicago Press. H. Elliot, Trans.

Lambert, H., G. Carder, and N. D'Cruze (2019). Given the cold shoulder: A review of the scientific literature for evidence of reptile sentience. *Animals 9*(10), 821.

Lambert, H., A. Elwin, and N. D'Cruze (2021). Wouldn't hurt a fly? A review of insect cognition and sentience in relation to their use as food and feed. *Applied Animal Behaviour Science 243*, 105432.

Laughlin, S. B. and M. Weckström (1993). Fast and slow photoreceptors—a comparative study of the functional diversity of coding and conductances in the Diptera. *Journal of Comparative Physiology A 172*(5), 593–609.

Lea, S. E. and W. H. Dittrich (2000). What do birds see in moving video images. *Picture perception in animals*, 143–180.

Learmonth, M. J. (2020). The matter of non-avian reptile sentience, and why it "matters" to them: A conceptual, ethical and scientific review. *Animals 10*(5), 901.

LeDoux, J. (2019). *The Deep History of Ourselves: The Four-Billion-Year Story of How We Got Conscious Brains*. New York: Viking.

LeDoux, J. E. (2017a). Emotional consciousness in animals: The case of fearful feelings. *Talk at NYU 2017 Animal Consciousness Conference*. `https://www.youtube.com/playlist?list=PLY_s7b9LrR8UCcPqL59XuIII68ALjgLt7` [Accessed on 12/03/2021].

LeDoux, J. E. (2017b). Semantics, surplus meaning, and the science of fear. *Trends in Cognitive Sciences 21*(5), 303–306.

LeDoux, J. E. (2022). As soon as there was life, there was danger: the deep history of survival behaviours and the shallower history of consciousness. *Philosophical Transactions of the Royal Society B 377*(1844), 20210292.

Lee, A. Y. (forthcoming). Degrees of Consciousness. *Noûs*.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific philosophical quarterly 64*(4), 354–361.

Levins, R. and R. C. Lewontin (1985). *The Dialectical Biologist*. Cambridge, MA: Harvard University Press.

Levy, D. and P. Glimcher (2016). Common value representation—A neuroeconomics perspective. In T. Brosch and D. Sander (Eds.), *Handbook of Value: Perspectives from Economics, Neuroscience, Philosophy, Psychology and Sociology*, pp. 85–118. Oxford: Oxford University Press.

Levy, D. J. and P. W. Glimcher (2012). The root of all value: a neural common currency for choice. *Current opinion in neurobiology 22*(6), 1027–1038.

Levy, G. and B. Hochner (2017). Embodied organization of Octopus vulgaris morphology, vision, and locomotion. *Frontiers in Physiology 8*, 164.

Lewontin, R. (1989). The evolution of cognition: questions we will never answer. In D. Osherson (Ed.), *An Invitation to Cognitive Science, Vol. 3*. Cambridge, MA: MIT Press.

Lewontin, R. and R. Levins (1997). Organism and environment. *CNS 8*(2), 95–98.

Lewontin, R. C. (1979). Sociobiology as an adaptationist program. *Behavioral science 24*(1), 5–14.

Lewontin, R. C. (1981). On constraints and adaptation. *Behavioral and Brain Sciences 4*(2), 244–245.

Lewontin, R. C. (1985). The organism as the subject and object of evolution. In R. Levins and R. C. Lewontin (Eds.), *The Dialectical Biologist*. Cambridge, MA: Harvard University Press.

Lichtenstein, S. and P. Slovic (2006). *The Construction of Preference*. Cambridge: Cambridge University Press.

Liljenström, H. and P. Århem (2008). *Consciousness Transitions: Phylogenetic, Ontogenetic and Physiological Aspects*. Amsterdam: Elsevier.

Lorenz, K. (1950). The comparative method in studying innate behavior patterns. In *Physiological Mechanisms in Animal Behavior. (Society's Symposium IV.)*., pp. 221–268. Oxford: Academic Press.

Lorenz, K. (1981). *The Foundations of Ethology*. Cham: Springer.

Lorenz, K. (1997). *The Natural Science of the Human Species. An Introduction to Comparative Behavioral Research. The "Russian Manuscript"*. Cambridge, MA: MIT Press.

Lyon, P. (2006). *The Agent in the Organism: Towards a Biogenic Theory of Cognition*. Ph.D. thesis, The Australian National University.

Machery, E. and J. Sytsma (2011). Robot pains and corporate feelings. *The Philosophers' Magazine* (52), 78–82.

Macphail, E. (1998). *The Evolution of Consciousness*. Oxford: Oxford University Press.

Maloof, A. C., S. M. Porter, J. L. Moore, F. Ö. Dudás, S. A. Bowring, J. A. Higgins, D. A. Fike, and M. P. Eddy (2010). The earliest Cambrian record of animals and ocean geochemical change. *GSA Bulletin 122*(11-12), 1731–1774.

Mangel, M. and C. W. Clark (1986). Towards a unifield foraging theory. *Ecology 67*(5), 1127–1138.

Manus, M. B. (2018). Evolutionary mismatch. *Evolution, Medicine, and Public Health 2018*(1), 190–191.

Margulis, L. (2001). The conscious cell. *Annals of the New York Academy of Sciences 929*(1), 55–70.

Marshall, N. J. and J. B. Messenger (1996). Colour-blind camouflage. *Nature 382*(6590), 408–409.

Marshall, R. and P. Godfrey-Smith (2014). Philosophy of biology: Peter Godfrey-Smith interviewed by Richard Marshall. *3:16*. `https://www.3-16am.co.uk/articles/philosophy-of-biology` [Accessed on 02/01/2022].

Maruzzo, D. and F. Bortolin (2013). Arthropod regeneration. In A. Minelli, G. Boxshall, and G. Fusco (Eds.), *Arthropod Biology and Evolution*, pp. 149–169. Cham: Springer.

Mascetti, G. G. (2016). Unihemispheric sleep and asymmetrical sleep: behavioral, neurophysiological, and functional perspectives. *Nature and Science of Sleep 8*, 221.

Mashour, G. A. and M. T. Alkire (2013). Evolution of consciousness: phylogeny, ontogeny, and emergence from general anesthesia. *Proceedings of the National Academy of Sciences 110*(Supplement 2), 10357–10364.

Massimini, M., F. Ferrarelli, R. Huber, S. K. Esser, H. Singh, and G. Tononi (2005). Breakdown of cortical effective connectivity during sleep. *Science 309*(5744), 2228–2232.

Mather, J. (2019). What is in an octopus's mind? *Animal Sentience 26*(1).

Mather, J. (2021a). The case for octopus consciousness: Unity. *NeuroSci 2*(4), 405–415.

Mather, J. (2021b). Octopus consciousness: the role of perceptual richness. *NeuroSci 2*(3), 276–290.

Matthewson, J. (2020). Does proper function come in degrees? *Biology & Philosophy 35*(4), 1–18.

Matthewson, J. and P. E. Griffiths (2017). Biological criteria of disease: Four ways of going wrong. *Journal of Medicine and Philosophy 42*(4), 447–466.

Maturana, H. R. and F. J. Varela (1980). Autopoiesis and Cognition: The Realization of the Living.

Maynard Smith, J. (1952). The importance of the nervous system in the evolution of animal flight. *Evolution 6*(1), 127–129.

Maynard Smith, J. (1987). Evolutionary progress and levels of selection. In J. Dupré (Ed.), *The Latest on the Best: Essays on Evolution and Optimality*, pp. 219–230. Cambridge, MA: MIT Press.

Mayr, E. (1988). *Toward a New Philosophy of Biology: Observations of an Evolutionist*. Cambridge, MA: Harvard University Press.

Mayr, E. (1994). Typological versus population thinking. *Conceptual issues in evolutionary biology*, 157–160.

McCleery, R. (1977). On satiation curves. *Animal Behaviour 25*, 1005–1015.

McFarland, D. and R. Sibly (1975). The behavioural final common path. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences 270*(907), 265–293.

McNamara, J. M. (1997). Optimal life histories for structured populations in fluctuating environments. *Theoretical Population Biology 51*(2), 94–108.

McNamara, J. M. and A. I. Houston (1986). The common currency for behavioral decisions. *The American Naturalist 127*(3), 358–378.

McNamara, J. M. and A. I. Houston (1996). State-dependent life histories. *Nature 380*(6571), 215–221.

McNamara, J. M. and A. I. Houston (2008). Optimal annual routines: behaviour in the context of physiology and ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences 363*(1490), 301–319.

McNamara, J. M. and A. I. Houston (2009). Integrating function and mechanism. *Trends in Ecology & Evolution 24*(12), 670–675.

Melis, A. P., J. Call, and M. Tomasello (2006). Chimpanzees (Pan troglodytes) conceal visual and auditory information from others. *Journal of Comparative Psychology 120*(2), 154.

Meltzoff, A. N. (2007). 'Like me': a foundation for social cognition. *Developmental Science 10*(1), 126–134.

Mendl, M., O. H. P. Burman, and E. S. Paul (2010). An integrative and functional framework for the study of animal emotion and mood. *Proceedings of the Royal Society B: Biological Sciences 277*(1696), 2895–2904.

Merker, B. (2005). The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution. *Consciousness and Cognition 14*(1), 89–114.

Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences 30*(1), 63–81.

Merker, B., K. Williford, and D. Rudrauf (2021). The Integrated Information Theory of consciousness: A case of mistaken identity. *Behavioral and Brain Sciences*, 1–72.

Merker, B., K. Williford, and D. Rudrauf (2022). The integrated information theory of consciousness: Unmasked and identified. *Behavioral and Brain Sciences 45*.

Messenger, J. B. (1977). Evidence that Octopus is colour blind. *Journal of Experimental Biology 70*(1), 49–55.

Metz, J. A., R. M. Nisbet, and S. A. Geritz (1992). How should we define 'fitness' for general ecological scenarios? *Trends in Ecology & Evolution 7*(6), 198–202.

Meyer, A. (1895). How was Wallace led to the discovery of natural selection? *Nature 52*(1348), 415–415.

Mikhalevich, I. and R. Powell (2020). Minds without spines: Evolutionarily inclusive animal ethics. *Animal Sentience 29*(1).

Millikan, R. G. (1995). *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

Millikan, R. G. (2002). Biofunctions: Two paradigms. In R. Cummins, A. Ariew, and M. Perlman (Eds.), *Functions: New Readings in the Philosophy of Psychology and Biology*, pp. 113–143. Oxford: Oxford University Press.

Misof, B., S. Liu, K. Meusemann, R. S. Peters, A. Donath, C. Mayer, P. B. Frandsen, J. Ware, T. Flouri, R. G. Beutel, et al. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science 346*(6210), 763–767.

Mitoh, S. and Y. Yusa (2021). Extreme autotomy and whole-body regeneration in photosynthetic sea slugs. *Current Biology 31*(5), R233–R234.

Morbeck, M., A. Galloway, and A. Zihlman (1997). *The Evolving Female: A Life-History Perspective*. Princeton, NJ: Princeton University Press.

Morrison, R. and D. Reiss (2018). Precocious development of self-awareness in dolphins. *PLoS One 13*(1), e0189813.

Müller, G. B. (2017). Why an extended evolutionary synthesis is necessary. *Interface Focus 7*(5), 20170015.

Murphy, D. (2020). Concepts of Disease and Health. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020

ed.). https://plato.stanford.edu/archives/sum2020/entries/health-disease/ [Accessed 20/07/2020].

Nagasako, E. M., A. L. Oaklander, and R. H. Dworkin (2003). Congenital insensitivity to pain: an update. *Pain 101*(3), 213–219.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review 83*(4), 435–450.

Nagel, T. (1986). *The View From Nowhere*. Oxford: Oxford University Press.

Nagel, T. (2012). *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False*. Oxford: Oxford University Press.

Narins, P. M., D. S. Grabul, K. K. Soma, P. Gaucher, and W. Hödl (2005). Cross-modal integration in a dart-poison frog. *Proceedings of the National Academy of sciences 102*(7), 2425–2429.

Nesher, N., G. Levy, F. W. Grasso, and B. Hochner (2014). Self-recognition mechanism between skin and suckers prevents octopus arms from interfering with each other. *Current Biology 24*(11), 1271–1275.

Nesse, R. M. (2007). The importance of evolution for medicine. In W. R. Trevathan, J. J. McKenna, and E. O. Smith (Eds.), *Evolutionary Medicine, Second Edition*, pp. 416–432. New York: Oxford University Press.

Nesse, R. M. and G. C. Williams (1995). *Evolution and Healing: The New Science of Darwinian Medicine*. London: Weidenfeld & Nicolson.

Neto, C. (2020). When imprecision is a good thing, or how imprecise concepts facilitate integration in biology. *Biology & Philosophy 35*(6), 1–21.

Nettle, D. and W. E. Frankenhuis (2020). Life-history theory in psychology and evolutionary biology: one research programme or two? *Philosophical Transactions of the Royal Society B 375*(1803), 20190490.

Nichols, S. and S. P. Stich (2003). *Mindreading: an integrated account of pretence, self-awareness, and understanding other minds.* New York: Oxford University Press.

Nicholson, D. J. (2014). The return of the organism as a fundamental explanatory concept in biology. *Philosophy Compass 9*(5), 347–359.

Nobel Prize Outreach (2021). Press release: The Nobel Prize in Physiology or Medicine 1973. *NobelPrize.org*. `https://www.nobelprize.org/prizes/medicine/1973/press-release/` [Accessed on 30/6/2021].

Noble, D. (2015). Evolution beyond Neo-Darwinism: a new conceptual framework. *Journal of Experimental Biology 218*(1), 7–13.

Nordenfelt, L. (1995). *On the Nature of Health: An Action-Theoretic Approach*. Dortrecht: Kluwer.

O Connell, S. and R. Dunbar (2003). A test for comprehension of false belief in chimpanzees. *Evolution and Cognition 9*(2), 131–140.

Okasha, S. (2018). *Agents and Goals in Evolution*. Oxford: Oxford University Press.

Ortega, L. J., K. Stoppa, O. Güntürkün, and N. F. Troje (2008). Limits of intraocular and interocular transfer in pigeons. *Behavioural Brain Research 193*(1), 69–78.

Ostojić, L., E. W. Legg, K. F. Brecht, F. Lange, C. Deininger, M. Mendl, and N. S. Clayton (2017). Current desires of conspecific observers affect cache-protection strategies in California scrub-jays and Eurasian jays. *Current Biology 27*(2), R51–R53.

Osvath, M. and M. Sima (2014). Sub-adult ravens synchronize their play: a case of emotional contagion. *Animal Behavior and Cognition 1*(2), 197–205.

Paley, W. (1802). *Natural Theology: or, Evidences of the Existence and Attributes of the Deity*. London: J. Faulder.

Palombo, D. J., C. Alain, H. Söderlund, W. Khuu, and B. Levine (2015). Severely deficient autobiographical memory (SDAM) in healthy adults: A new mnemonic syndrome. *Neuropsychologia 72*, 105–118.

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford: Oxford University Press.

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition 14*(1), 30–80.

Panksepp, J. (2010). Affective consciousness in animals: perspectives on dimensional and primary process emotion approaches. *Proceedings of the Royal Society B: Biological Sciences 277*(1696), 2905–2907.

Panksepp, J. (2011). Cross-species affective neuroscience decoding of the primal affective experiences of humans and related animals. *PloS One 6*(9), e21236.

Parkhaev, P. Y. (2007). The Cambrian 'basement'of gastropod evolution. *Geological Society, London, Special Publications 286*(1), 415–421.

Pearce, J. M., G. R. Esber, D. N. George, and M. Haselgrove (2008). The nature of discrimination learning in pigeons. *Learning & Behavior 36*(3), 188–199.

Pearson, J. (2019). The human imagination: the cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience 20*(10), 624–634.

Pearson, J. M., K. K. Watson, and M. L. Platt (2014). Decision making: the neuroethological turn. *Neuron 82*(5), 950–965.

Pepperberg, I. M., S. E. Garcia, E. C. Jackson, and S. Marconi (1995). Mirror use by African grey parrots (Psittacus erithacus). *Journal of Comparative Psychology 109*(2), 182.

Peressini, A. (2014). Blurring two conceptions of subjective experience: Folk versus philosophical phenomenality. *Philosophical Psychology 27*(6), 862–889.

Perry, C. J., L. Baciadonna, and L. Chittka (2016). Unexpected rewards induce dopamine-dependent positive emotion–like state changes in bumblebees. *Science 353*(6307), 1529–1531.

Perry, C. J., A. B. Barron, and K. Cheng (2013). Invertebrate learning and cognition: relating phenomena to neural substrate. *Wiley Interdisciplinary Reviews: Cognitive Science 4*(5), 561–582.

Peterson, K. J., J. A. Cotton, J. G. Gehling, and D. Pisani (2008). The Ediacaran emergence of bilaterians: congruence between the genetic and the geological fossil records. *Philosophical Transactions of the Royal Society B: Biological Sciences 363*(1496), 1435–1443.

Phillips, J., W. Buckwalter, F. Cushman, O. Friedman, A. Martin, J. Turri, L. Santos, and J. Knobe (2020). Knowledge before Belief. *Behavioral and Brain Sciences*, 1–37.

Pianka, E. R. and W. S. Parker (1975). Age-specific reproductive tactics. *The American Naturalist 109*(968), 453–464.

Pirotta, E., M. Mangel, D. P. Costa, B. Mate, J. A. Goldbogen, D. M. Palacios, L. A. Hückstädt, E. A. McHuron, L. Schwarz, and L. New (2018). A dynamic state model of migratory behavior and physiology to assess the consequences of environmental variation and anthropogenic disturbance on marine vertebrates. *The American Naturalist 191*(2), E40–E56.

Pittendrigh, C. S. (1958). Adaptation, natural selection, and behavior. In A. Roe and G. G. Simpson (Eds.), *Behavior and Evolution*. New Haven: Yale University Press.

Plotnik, J. M., F. B. De Waal, and D. Reiss (2006). Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences 103*(45), 17053–17057.

Plutchik, R. (1962). *The Emotions: Facts, Theories and a New Model*. New York: Random House.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik and H. Kellerman (Eds.), *Emotion: Theory, Research, and Experience*, pp. 3–33. Cambridge, MA: Academic Press.

Potier, S., M. Lieuvin, M. Pfaff, and A. Kelber (2020). How fast can raptors see? *Journal of Experimental Biology 223*(1), jeb209031.

Premack, D. and G. Woodruff (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences 1*(4), 515–526.

Prinz, J. (2000). A neurofunctional theory of visual consciousness. *Consciousness and Cognition 9*(2), 243–259.

Prinz, J. J. (2012). *The Conscious Brain: How Attention Engenders Experience*. Oxford: Oxford University Press.

Prior, H., A. Schwarz, and O. Güntürkün (2008). Mirror-induced behavior in the magpie (Pica pica): evidence of self-recognition. *PLoS Biology 6*(8), e202.

Railo, H. and M. Hurme (2021). Is the primary visual cortex necessary for blindsight-like behavior? Review of transcranial magnetic stimulation studies in neurologically healthy individuals. *Neuroscience & Biobehavioral Reviews 127*, 353–364.

Rattenborg, N. C., B. Voirin, S. M. Cruz, R. Tisdale, G. Dell'Omo, H.-P. Lipp, M. Wikelski, and A. L. Vyssotski (2016). Evidence that birds sleep in mid-flight. *Nature Communications 7*(1), 1–9.

Reber, A. S. (2019). *The First Minds: Caterpillars, Karyotes, and Consciousness*. Oxford: Oxford University Press.

Rich, A. N., J. L. Bradshaw, and J. B. Mattingley (2005). A systematic, large-scale study of synaesthesia: implications for the role of early experience in lexical-colour associations. *Cognition 98*(1), 53–84.

Richards, R. J. (1992). *The Meaning of Evolution: The Morphological Construction and Ideological Reconstruction of Darwin's Theory*. Chicago: University of Chicago Press.

Ristau, C. A. (1992). Cognitive ethology: Past, present and speculations on the future. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Volume 1992, pp. 125–136. Philosophy of Science Association.

Ro, T., D. Shelton, O. L. Lee, and E. Chang (2004). Extrageniculate mediation of unconscious vision in transcranial magnetic stimulation-induced blindsight. *Proceedings of the National Academy of Sciences 101*(26), 9933–9935.

Rodenburg, T., A. Buitenhuis, B. Ask, K. Uitdehaag, P. Koene, J. Van der Poel, and H. Bovenhuis (2003). Heritability of feather pecking and open-field response of laying hens at two different ages. *Poultry science 82*(6), 861–867.

Roff, D. A. (1992). *Evolution of Life Histories: Theory and Analysis*. New York: Chapman and Hall.

Roff, D. A. (2002). *Life History Evolution*. New York: W. H. Freeman.

Rogers, L. J., G. Vallortigara, and R. J. Andrew (2013). *Divided Brains: The Biology and Behaviour of Brain Asymmetries*. Cambridge: Cambridge University Press.

Rolls, E. T. (1999). *The Brain and Emotion*. Oxford: Oxford University Press.

Romanes, G. J. (1883). *Mental Evolution in Animals*. London: Kegan Paul, Trench, & Co.

Ross, D. (2020). Addiction is socially engineered exploitation of natural biological vulnerability. *Behavioural Brain Research 386*, 112598.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review 110*(1), 145.

Russell, J. A. and L. F. Barrett (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of Personality and Social Psychology 76*(5), 805.

Russell, J. A. and B. Fehr (1987). Relativity in the perception of emotion in facial expressions. *Journal of Experimental Psychology: General 116*(3), 223.

Samuelson, L. and J. M. Swinkels (2006). Information, evolution and utility. *Theoretical Economics 1*(1), 119–142.

Saxe, R. and N. Kanwisher (2003). People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". *Neuroimage 19*(4), 1835–1842.

Schechter, E. (2018). *Self-Consciousness and "Split" Brains: The Minds' I*. Oxford: Oxford University Press.

Schnell, A. K. and N. S. Clayton (2021). Cephalopods: Ambassadors for rethinking cognition. *Biochemical and Biophysical Research Communications 564*, 27–36.

Schnell, A. K., R. T. Hanlon, A. Benkada, and C. Jozet-Alves (2016). Lateralization of eye use in cuttlefish: opposite direction for anti-predatory and predatory behaviors. *Frontiers in Physiology 7*, 620.

Schukraft, J. (2020). Does critical flicker-fusion frequency track the subjective experience of time? *Rethink Priorities*. https://rethinkpriorities.org/publications/does-critical-flicker-fusion-frequency-track-the-subjective-experience-of-time [Accessed on 06/05/2022].

Schwartz, S. K., W. E. Wagner Jr, and E. A. Hebets (2016). Males can benefit from sexual cannibalism facilitated by self-sacrifice. *Current Biology 26*(20), 2794–2799.

Sebo, J. (2018). The moral problem of other minds. *The Harvard Review of Philosophy 25*, 51–70.

Sheets-Johnstone, M. (1999). *The Primacy of Movement*. Amsterdam: John Benjamins Pub.

Shen, B., L. Dong, S. Xiao, and M. Kowalewski (2008). The Avalon explosion: evolution of Ediacara morphospace. *Science 319*(5859), 81–84.

Shermer, M. (2002). *In Darwin's Shadow: The Life and Science of Alfred Russel Wallace: A Biographical Study on the Psychology of History*. Oxford: Oxford University Press.

Sherrington, C. S. (1906). *The Integrative Action of the Nervous System*. New Haven, CT: Yale University Press.

Shizgal, P. and K. Conover (1996). On the neural computation of utility. *Current Directions in Psychological Science 5*(2), 37–43.

Sneddon, L. U., J. Lopez-Luna, D. C. Wolfenden, M. C. Leach, A. M. Valentim, P. J. Steenbergen, N. Bardine, A. D. Currie, D. M. Broom, and C. Brown (2018). Fish sentience denial: Muddying the waters. *Animal Sentience 3*(21), 1.

Solms, M. (2021). *The Hidden Spring: A Journey to the Source of Consciousness*. New York: WW Norton & Company.

Sovrano, V. A., A. Bisazza, and G. Vallortigara (2001). Lateralization of response to social stimuli in fishes: a comparison between different methods and species. *Physiology & Behavior 74*(1-2), 237–244.

Spencer, H. (1855). *Principles of Psychology*. London: Longman, Brown and Green.

Spencer, H. (1870). *First Principles*. London: Williams and Norgate.

Sprecher, S. G., A. Cardona, and V. Hartenstein (2011). The Drosophila larval visual system: high-resolution analysis of a simple visual neuropil. *Developmental Biology 358*(1), 33–43.

Spurrett, D. (2014). Philosophers should be interested in 'common currency'claims in the cognitive and behavioural sciences. *South African Journal of Philosophy 33*(2), 211–221.

Spurrett, D. (2015). The natural history of desire. *South African Journal of Philosophy 34*(3), 304–313.

Spurrett, D. (2020). The descent of preferences. *The British Journal for the Philosophy of Science*.

Stearns, S. C. (1992). *The Evolution of Life Histories*. Oxford: Oxford University Press.

Sterelny, K. (1997). Where does thinking come from? A commentary on Peter Godfrey-Smith's Complexity and the Function of Mind in Nature. *Biology & Philosophy 12*(4), 551–566.

Sterelny, K. (2003). *Thought in a Hostile World*. Oxford: Blackwell.

Sterelny, K. and P. E. Griffiths (1999). *Sex and Death: An Introduction to Philosophy of Biology*. Chicago: University of Chicago press.

Stoltzfus, A. (2019). Understanding bias in the introduction of variation as an evolutionary cause. In T. Uller and K. N. Laland (Eds.), *Evolutionary Causation: Biological and Philosophical Reflections*, pp. 29–61. Cambridge, MA:MIT Press.

Suzuki, Y., J. Chou, S. L. Garvey, V. R. Wang, and K. O. Yanes (2019). Evolution and regulation of limb regeneration in arthropods. *Evo-Devo: Non-model Species in Cell and Developmental Biology*, 419–454.

Sytsma, J. (2010). Dennett's theory of the folk theory of consciousness. *Journal of Consciousness Studies 17*(3-4), 107–130.

Sytsma, J. (2012). Revisiting the valence account. *Philosophical Topics 40*(2), 179–198.

Sytsma, J. and E. Machery (2010). Two conceptions of subjective experience. *Philosophical Studies 151*(2), 299–327.

Sytsma, J. and E. Machery (2012). On the relevance of folk intuitions: A commentary on Talbot. *Consciousness and Cognition 21*(2), 654–660.

Sytsma, J. and E. Ozdemir (2019). No problem: Evidence that the concept of phenomenal consciousness is not widespread. *Journal of Consciousness Studies 26*(9-10), 241–256.

Sytsma, J. M. and E. Machery (2009). How to study folk intuitions about phenomenal consciousness1. *Philosophical Psychology 22*(1), 21–35.

Talbot, B. (2012). The irrelevance of dispositions and difficulty to intuitions about the "hard problem" of consciousness: A response to Sytsma, Machery, and Huebner. *Consciousness and Cognition 21*(2), 661–666.

Tanaka, M. M., P. Godfrey-Smith, and B. Kerr (2020). The dual landscape model of adaptation and niche construction. *Philosophy of Science 87*(3), 478–498.

Thompson, E. (2007). *Mind in Life*. Cambridge, MA: Harvard University Press.

Thompson, E. (2011). Reply to commentaries. *Journal of Consciousness Studies 18*(5-6), 176–223.

Thompson, E. (2022). Could all life be sentient? *Journal of Consciousness Studies 29*(3-4), 229–265.

Thompson, N. S. (1986a). Deception and the concept of behavioral design. In R. W. Mitchell and N. Thompson (Eds.), *Deception*, pp. 53–66. Albany, NY: SUNY Press.

Thompson, N. S. (1986b). Ethology and the birth of comparative teleonomy. In R. Campan and R. Zayan (Eds.), *Relevance of Models and Theory in Ethology*, pp. 13–23. Toulouse: Privat, I. E. C.

Thrush, D. (1973). Congenital insensitivity to pain. *Brain 96*(2), 369–386.

Todd, J. T. and E. K. Morris (1986). The early research of John B. Watson: Before the behavioral revolution. *The Behavior Analyst 9*(1), 71–88.

Tolman, E. (1923). The nature of instinct. *Psychological Bulletin 20*(4), 200–218.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience 5*(1), 42.

Tononi, G. (2005). Consciousness, information integration, and the brain. *Progress in Brain Research 150*, 109–126.

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *The Biological Bulletin 215*(3), 216–242.

Tononi, G. (2010). Information integration: its relevance to brain function and consciousness. *Archives Italiennes de Biologie 148*(3), 299–322.

Tononi, G. (2012a). The integrated information theory of consciousness: an updated account. *Archives Italiennes de Biologie 150*(2/3), 56–90.

Tononi, G. (2012b). *Phi: A Voyage from the Brain to the Soul*. New York: Pantheon Books.

Tononi, G., M. Boly, M. Grasso, J. Hendren, B. E. Juel, W. G. Mayner, W. Marshall, and C. Koch (2022). IIT, half masked and half disfigured. *Behavioral and Brain Sciences 45*, e60.

Tononi, G., M. Boly, M. Massimini, and C. Koch (2016). Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience 17*(7), 450–461.

Tononi, G. and G. M. Edelman (1998). Consciousness and complexity. *Science 282*(5395), 1846–1851.

Tononi, G. and C. Koch (2015). Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences 370*(1668), 20140167.

Travers, E., C. D. Frith, and N. Shea (2018). Learning rapidly about the relevance of visual cues requires conscious awareness. *Quarterly Journal of Experimental Psychology 71*(8), 1698–1713.

Trestman, M. (2013). The Cambrian explosion and the origins of embodied cognition. *Biological Theory 8*(1), 80–92.

Trestman, M. (2017). Minds and bodies in animal evolution. In K. Andrews and J. Beck (Eds.), *The Routledge Handbook of Animals Minds*, pp. 206–215. New York: Routledge.

Trewavas, A., F. Baluška, S. Mancuso, and P. Calvo (2020). Consciousness facilitates plant behavior. *Trends in Plant Science 25*(3), 216–217.

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.

Tye, M. (2016). *Tense Bees and Shell-Shocked Crabs: Are Animals Conscious?* Oxford: Oxford University Press.

Tye, M. (2021). Qualia. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021 ed.). https://plato.stanford.edu/archives/fall2021/entries/qualia/ [Accessed on 02/01/2022].

Vallortigara, G. (2000). Comparative neuropsychology of the dual brain: a stroll through animals' left and right perceptual worlds. *Brain and Language 73*(2), 189–219.

Vallortigara, G. and L. J. Rogers (2020). A function for the bicameral mind. *Cortex 124*, 274–285.

Veit, W. (2019). Evolution of multicellularity: cheating done right. *Biology & Philosophy 34*(3), 34.

Veit, W. (2021a). Agential thinking. *Synthese 199*(5), 13393–13419.

Veit, W. (2021b). The evolution of knowledge during the Cambrian explosion. *Behavioral and Brain Sciences 44*, e47.

Veit, W. (2021c). Experimental philosophy of medicine and the concepts of health and disease. *Theoretical Medicine and Bioethics 42*(3), 169–186.

Veit, W. (2021d). Samir Okasha's philosophy. *Lato Sensu: revue de la Société de philosophie des sciences 8*(3), 1–8.

Veit, W. (2022a). Complexity and the evolution of consciousness. *Biological Theory*. https://doi.org/10.1007/s13752-022-00407-z.

Veit, W. (2022b). Consciousness, complexity, and evolution. *Behavioral and Brain Sciences 45*, e61.

Veit, W. (2022c). The origins of consciousness or the war of the five dimensions. *Biological Theory*. https://doi.org/10.1007/s13752-022-00408-y.

Veit, W. (2022d). Review of Peter Godfrey-Smith's Metazoa: Animal Minds and the Birth of Consciousness. *Philosophy of Science 89*(3), 658–660.

Veit, W. (2022e). Scaffolding natural selection. *Biological Theory 17*(2), 163–180.

Veit, W. (2022f). Towards a comparative study of animal consciousness. *Biological Theory*. https://doi.org/10.1007/s13752-022-00409-x.

Veit, W. (forthcoming). Health, Consciousness, and the Evolution of Subjects. *Synthese*.

Veit, W. and H. Browning (2020a). Perspectival pluralism for animal welfare. *European Journal for Philosophy of Science 11*(1), 1–14.

Veit, W. and H. Browning (2020b). Two kinds of conceptual engineering. *Preprint*. http://philsci-archive.pitt.edu/17452/.

Veit, W. and H. Browning (2021). Extending animal welfare science to include wild animals. *Animal Sentience 7*(20), 1–4.

Veit, W. and H. Browning (forthcoming). Pathological complexity and the evolution of sex differences. *Behavioral and Brain Sciences*.

Veit, W. and B. Huebner (2020). Drawing the boundaries of animal sentience. *Animal Sentience 29*(13).

Veit, W. and M. Ney (2021). Metaphors in arts and science. *European Journal for Philosophy of Science 11*(2), 1–24.

Velmans, M. (1991). Is human information processing conscious? *Behavioral and Brain Sciences 14*(4), 651–669.

Vonk, J. (2019). A fish eye view of the mirror test. *Learning & Behavior*, 1–2.

Waggoner, B. (2003). The Ediacaran biotas in space and time. *Integrative and Comparative Biology 43*(1), 104–113.

Walsh, D. M. (2015). *Organisms, Agency, and Evolution*. Cambridge: Cambridge University Press.

Walters, E. T. (2018). Nociceptive biology of molluscs and arthropods: evolutionary clues about functions and mechanisms potentially related to pain. *Frontiers in Physiology 9*, 1049.

Walters, E. T., T. J. Carew, and E. R. Kandel (1981). Associative learning in Aplysia: Evidence for conditioned fear in an invertebrate. *Science 211*(4481), 504–506.

Ward, J. (2013). Synesthesia. *Annual Review of Psychology 64*, 49–75.

Watkins, N. W. (2018). (A)phantasia and severely deficient autobiographical memory: Scientific and personal perspectives. *Cortex 105*, 41–52.

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

Watson, M. R., K. A. Akins, C. Spiker, L. Crawford, and J. T. Enns (2014). Synesthesia and learning: a critical review and novel theory. *Frontiers in Human Neuroscience 8*, 98.

Wellman, H. M. (2014). *Making Minds: How Theory of Mind Develops*. Oxford: Oxford University Press.

Wells, M. (1978). *Octopus: Physiology and Behaviour of an Advanced Invertebrate*. London: Chapman and Hall.

Whiten, A. (2013). Humans are not alone in computing how others see the world. *Animal Behaviour 86*(2), 213–221.

Wigglesworth, V. B. (1980). Do insects feel pain. *Antenna 4*, 8–9.

Wilkes, K. V. (1984). Is consciousness important? *British Journal for the Philosophy of Science*, 223–243.

Woodruff, M. L. (2017). Consciousness in teleosts: There is something it feels like to be a fish. *Animal Sentience 13*(1).

Wray, G. A. (2015). Molecular clocks and the early evolution of metazoan nervous systems. *Philosophical Transactions of the Royal Society B: Biological Sciences 370*(1684), 20150046.

Zeman, A., M. Dewar, and S. D. Sala (2015). Lives without imagery – Congenital aphantasia. *Cortex 73*, 378–380.

Zuk, M. (2016). Mates with benefits: when and how sexual cannibalism is adaptive. *Current Biology 26*(23), R1230–R1232.