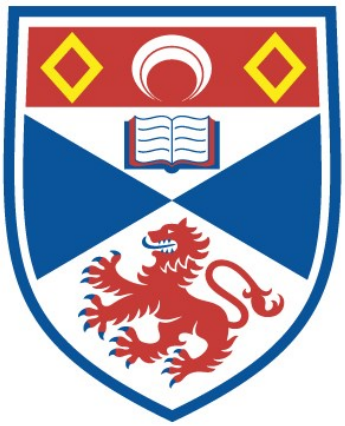
Productive being

The continuity between life and action

Mirko Alexander Prokop

A thesis submitted for the degree of MPhil
at the
University of St Andrews



2022

Identifier to use to cite or link to this thesis:

DOI: https://doi.org/10.17630/sta/321

Candidate's declaration

I, Mirko Alexander Prokop, do hereby certify that this thesis, submitted for the degree of MPhil, which is approximately 40,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in September 2020.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date 06.11.2022 Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of MPhil in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date 07/11/2022 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Mirko Alexander Prokop, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

Embargo on all of electronic copy for a period of 1 year on the following ground(s):

- Publication would preclude future publication

Supporting statement for electronic embargo request

I have a publishing contract pending.

Title and Abstract

- I agree to the title and abstract being published.

Date 06.11.2022 Signature of candidate

Date 07/11/2022 Signature of supervisor.

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Mirko Alexander Prokop, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date 06.11.2022 Signature of candidate

General Acknowledgements

I would like to express my sincere gratitude to Mike Wheeler for his excellent supervision, our inspiring conversations and the unwavering support as well as to Adrian Haddock for introducing me to Anscombe's work and for many challenging discussions. Many thanks also to Ezequiel Di Paolo and participants in the ENSO reading group for illuminating exchanges about enactivism and beyond. Finally, I am deeply grateful to my family, friends and loved ones for being there for me in times of trouble as well as joy.

Funding

This work was supported by the German Academic Exchange Service (DAAD) [Scholarship ID: 57503736] and the University of St Andrews (Accommodation Award, 2020/1).

Productive Being

The Continuity between Life and Action

Abstract

What is the relationship between life and agency? Are all agents necessarily living organisms? How are we to understand the phenomena of life and action? This thesis aims to answer these questions against the background of the autopoietic enactive approach in cognitive science. I develop a general conception of productive beings as systems capable of engaging in productive processes. A productive process is understood as a purposive, end-directed kind of self-movement whose cause or source lies in the subject which engages in the process. I argue that the capacity of productive systems to engage in productive processes is inextricably tied to their nature as materially precarious systems. In other words, the autonomous form of productive beings cannot be understood without reference to their materially fragile realisation. I then argue that, given an enactive conception of life and precariousness, productive systems must be understood as living systems. Since agents, on the developed view, are productive systems, I conclude that life is necessary for agency, from an enactive viewpoint. The understanding of this continuity between life and action is explored, throughout the thesis, from different perspectives. Apart from a careful examination of enactive concepts and proposals, the discussion engages Elizabeth Anscombe's work on agency, Michael Thompson's reflections on life-forms, Philippa Foot's notion of natural goodness, Hans Jonas' philosophy of biology, and Aristotle's conception of the relationship between form and matter.

CONTENTS

Introduction

In January 2022, after a month long journey, the James Webb Space Telescope reached its home at the Lagrange point L_2. Since then, JWST has been stunning the world with awe-inspiring images of our universe. One particularly exciting prospect is JWST's capacity to study exoplanets. Indeed, recent observations provide the first firm evidence of carbon dioxide in an exoplanet's atmosphere, WASP-39b.[1] Carbon dioxide, alongside methane, is a so-called 'biosignature', an indicator for the presence of life (Hall 2022).

But what is it a signature *of*? How do we know that an exoplanet is not deceiving us, that it's the real thing: *life*? Suppose one day the composition of some exoplanet is highly promising. We fly out there and take a look. We encounter a flat, slimy, turquoise disk beside a bubbling pool. Is it alive? Samples are taken, flown back and analysed meticulously. The thing doesn't have DNA, but it slowly expands and contracts. It seems to require no nutrients, but a drop of water per day maintains its colour, and it somehow produces oxygen – clearly a biosignature! But is it alive?

Admittedly, this scenario is rather fanciful, but this should not impair the following point. For surely *before* all this sampling, analysing and signature-collecting, some bottom-rank NASA officer would have boldly suggested, there and then, 'just poke it and see it if moves', and if it *had* moved, chances are next day's headline would have read *Alien Life Found*. As this brings out, there is a fine line between our recognising a thing's aliveness and its capacity to move and change in certain ways. Broadly speaking, *life* and *action* are conceptually entangled phenomena in that the recognition of latter is naively taken as an indication of the former: when a thing moves or changes in certain ways, we tend to deem it alive. But is any agent necessarily alive? Conversely, is any living thing capable of actions? When does 'moving and changing in certain ways' amount to an *action*? And what is *life*, anyway?

The relationship between life and action, into which these questions inquire, is our topic. Understanding this relationship has important implications for the explanation of other phenomena such as purposiveness, cognition and consciousness, particularly from the perspective of one approach in cognitive science, so-called 'autopoietic enactivism'. Like other enactive strands (see Ward et. al. 2017), autopoietic enactivism – henceforth simply *the enactive approach* or *enactivism* – stresses the importance of system-environment interactions and the role of the body for the explanation of cognitive phenomena (E. Thompson 2007, Di Paolo et. al. 2017, 2018). Moreover, it emphasises the continuity between life and mind, i.e. the idea that life and mind share a set of basic principles which are central to the understanding of both phenomena (Clark 2001: 118, E. Thompson 2007: Ch. 6; cf. Godfrey-Smith 1994, Wheeler 1997).

Until recently, enactivism did not explicitly thematise *agency*, i.e. a system's capacity to perform actions (Barandiaran et. al. 2009, Di Paolo et. al. 2017). From an enactive, life-mind continuity perspective, however, the standing of agency in relation to life bears on the understanding of cognition more generally. Some

[1] See JWST Transiting Exoplanet Community Early Release Science Team. (2022).

proponents of enactivism suggest that even minimal living systems, e.g. bacteria, satisfy the conditions for agency (Barandiaran et. al. 2009: 374-376). Nevertheless, the suggestion that life is *sufficient* for agency is paired with a reluctance to claim that life is *necessary* for agency or cognition (ibid.: 376; E. Thompson 2011). Given the development of ever more sophisticated intelligent systems, e.g. autonomous robots, self-driving cars, or deep learning programs, which are seemingly able to *do* things despite being lifeless, this reluctance is understandable. Sure, empirically speaking, most or perhaps all genuine agents we know are living organisms. But why think that an agent *must* be alive?

The ensuing pages wish to challenge the idea that agency can be detached from life. More precisely, I hope to show that, from an enactive perspective, life is necessary for agency. Since, as will emerge, cognition, enactively understood, depends on agency, this means that life remains the necessary breeding ground for both action and mind.

Chapter 1 examines the web of concepts, with an emphasis on agency, which underwrites the enactive approach and, against this background, motivates more clearly our guiding question: is life necessary for agency? In Chapter 2, we consider Elizabeth Anscombe's approach to agency, and related themes such as Michael Thompson's concept of a *life-form* and Philippa Foot's notion of *natural goodness*. Following Anscombe, I develop an abstract, general conception of what I call a *productive process* and, relatedly, of a *productive system*. These notions allow the articulation of two explanatory constraints which, I propose, any account of agency should meet. Finally, Chapter 3 investigates whether the enactive conception of agency satisfies these constraints. I will argue that it can. However, through analysing an exchange between Wheeler (2011) and E. Thompson (2011) in the context of recent enactive developments, I submit that, in order to do so, agency must remain conceptually tied to life. In that sense, life is necessary for agency.

Drawing on Hans Jonas' analysis of metabolism and Aristotle's reflections on the relationship between form and matter, my central argument rests on two claims. First, to be an agent and hence a productive system is to lead an inherently fragile, precarious kind of existence. Second, the precariousness of productive being cannot be grasped without an explicit reference to the biochemical materiality of living systems. As we shall put it, following David Charles' (2021) interpretation of Aristotle, productive beings are *inextricably enmattered*: their mode of being is conceptually dependent on material properties.

The dependence of agency on life is, I think, implicit in enactivism, but it takes effort to show, on conceptual grounds, why and in which way the approach is committed to it. Of course, the phenomena of life and agency are so hopelessly vast that I will not pretend, within a hundred odd pages, to write their relationship into stone. Our inquiry will be of a relatively limited 'enactive' scope. That said, though, my background aim is to build bridges between enactive and Anscombian themes from which, I believe, both approaches benefit. At any rate, *if* the enactive approach *and* my arguments are sound, the prospects of finding a disembodied, inorganic intelligent agent somewhere in the universe are dim. The fine line between our recognition of agency and life is not a prejudice, but an insight to be cherished.

1 The Enactive Approach

1.1 Origins

Although contemporary 'autopoietic' enactivism has moved beyond Humberto Maturana and Francisco Varela's (abbreviated as 'MV') original theory of autopoiesis, the label is adequate since the enactive approach remains rooted in MV's work. In this first part of Ch. 1, we introduce the general idea behind MV's approach, before discussing the concepts of autopoiesis and autonomy. Finally, we consider the idea of structural coupling, the backdrop for the current enactive understanding of the continuity between life and mind.

1.1.1 Life and Organisation: The Starting Point

Life, C. S. Lewis remarks, is one of 'the greater words' (1967: 269) with a perplexing variety of connotations. The sense of *life* which interests us is the biological one, which, as Lewis observes, is 'the prime source of [the word's] magic' (1967: 293). Life, understood biologically, captures what is common to all living things, i.e. organisms, and what distinguishes them from non-living things. It corresponds to what Aristotle (2016) meant by the Greek *psyche*, or the Scholastics (e.g. Aquinas 1912: 2ff.) by the Latin *anima* – words translated as 'soul', 'essence' or 'self', but whose root meaning connotes the idea of the defining principle of life (Capra and Luisi 2014: 5).

The biological sense of life underlies the guiding question of MV's work *Autopoiesis: The Organization of the Living*: 'What is common to all living systems that we qualify them as living?' (1973: 74f.). In reply, MV did not appeal to an irreducible life-force, as the Vitalist movement did, nor did they draw up a list of properties common to organisms, e.g. reproduction, metabolism, evolution etc.[2] Instead, they took an *organisational approach* towards determining the essential characteristic of living systems:

> It is our assumption that there is an organization that is common to all living systems, whichever the nature of their components. Since our subject is this organization, not the particular ways in which it may be realized, we shall not make distinctions between classes or types of living systems. (Maturana and Varela 1973: 76).

The assumption that the common feature of organisms is specifiable in organisational terms has two consequences for one's approach towards life. First, the material parts, e.g. molecules, cells, tissues etc. of which an organism consists are deemed ultimately irrelevant for understanding why it is alive (ibid.: 77f.). Of course, to determine its organisation, one must examine the physical make-up of an organism. But once this organisation is understood, its physical realization may be ignored without losing sight of what makes the organism alive. Thus, on an

[2] See Maturana's *Introduction* in Maturana and Varela (1980), M. Thompson (2008: Ch. 2) for why a list-definition is unsatisfactory.

organisational approach, the fact that organisms consist of such-and-such materials, e.g. DNA, carbon, cells, etc. is contingent. It doesn't affect the 'organisational essence' (Moreno and Mossio 2015: 7) in which their aliveness consists. Second, the assumption that there is one living organisation implies that the understanding of life itself imposes no *order* on living kinds, e.g. species. Different life forms can certainly be arranged in various ways, e.g. according to evolutionary history, traits, habitats, interbreeding barriers, and so on. But such arrangements are not necessary to understand why they all *live*. We need not know the relations between living kinds or how their diversity evolved, e.g. through reproduction, in order to understand what makes them alive.

Given such far-reaching implications, why did MV choose an organisational approach towards life? There seem to be two reasons. First, they took the most essential feature of living organisms to be that they *actively sustain themselves as unities*. Second, they thought we can explain this self-sustaining capacity by understanding the particular kind of organisation common to living systems (1973: 73-76). MV's rationale thus combines two simple ideas: there is one most essential feature X common to all living organisms and we can explain X in organisational terms. *Voilà, c'est la vie!* – let us spoil this simplicity by unpacking these ideas.

The 'most essential feature' of organisms is what MV called their *autonomy*, as 'revealed in the self-asserting capacity of living systems to maintain their identity through the active compensation of deformations' (1973: 73). The autonomy of organisms, their ability to sustain their unity through their own activity, seems indeed to be a striking property of life, recognizable even for someone who knows nothing of evolution, physiology, biochemistry, or genetics. It is what Jonas has in mind when observing that the use of the term 'self' is 'unavoidable in any description of the most elementary instance of life' (2016: 55). It also underlies Kant's description of living organisms as self-organising beings or 'natural ends' (2000: §64-5) (more on this later) and Aristotle's (2016) view of the soul as a principle of self-movement in living things. Autonomy captures the intuitive, yet puzzling observation that organisms are able to sustain and move them*selves*, to *self*-organise, to exist thanks to their *own* activity. Thus, when MV announce that their objective is 'to understand the organisation of living systems in relation to their unitary character [i.e. autonomy]' (1973: 75), they align themselves with a long tradition of seeking to understand life in view of the self-unifying, endogenous activity of organisms.

Fixating autonomy as the essential starting point for the understanding of life invites an organisational approach. For the abovementioned implications of such an approach nicely 'track' our intuitive grasp of autonomy. That is, autonomy seems to belong to living things irrespective of their physical make-up or their position in evolutionary history or an ecosystem. Despite the bewildering variety of forms that organisms can take, we nonetheless discern in each of them a striking capacity to actively sustain itself as a unity. Aristotle certainly recognised the importance of this capacity for the understanding of life despite providing dubious physiological theories of its realisation (Ross 1964: Ch. IV). Kant also found life's self-organising capacity so central that he held on to it as an essential characteristic of our understanding of organisms despite contending that a naturalistic explanation of it

was out of human reach (Kant 2000: §75). With this provisional harmony between our intuitive grasp of autonomy and the implications of an organisational approach in mind, let's take a closer look at MV's proposal.

1.1.2 Autopoiesis: The Living Organisation

To vindicate their twofold rationale, MV had to explain autonomy in organisational terms. To this end, they proposed the concept of *autopoiesis*, meaning literally 'self-production' (from Greek *autós* = self and *poíēsis* = to make/create/produce). Organisms, in their view, are 'autopoietic machines':

> An autopoietic machine is a machine organised (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces the components which:
> (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and
> (ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as such a network. (Maturana and Varela 1973: 78f. Emphasis removed).

The term 'machine' is misleading insofar as it connotes lifeless, purposefully made artefacts, e.g. a steam engine. MV use the term interchangeably with 'system', i.e. an entity whose identity is defined by its organisation irrespective of its material parts (ibid.: 77). Following other commentators (E. Thompson 2007: 99f., Di Paolo 2005), I will speak of autopoietic 'systems', not 'machines'.

The living cell is the best understood and simplest case of an autopoietic system. In the cell, (i) biochemical processes, e.g. molecular reactions, continually reproduce the components, e.g. molecules, which, through their causal interplay, make these processes possible, thus forming a self-sustaining network of processes of production. Furthermore, (ii) this network is topologically bounded by generating a membrane which, apart from preserving the intracellular milieu in which the constituent processes can occur, distinguishes the network as a concrete unity in space. Thus, a cell is a self-producing and self-distinguishing network of processes in the biochemical domain (E. Thompson 2007: 44).

The self-distinction element of autopoiesis should not be confused with self-*isolation*. An autopoietic system relies on the constant interchange of material and energy across its boundary to reproduce the component processes which realise its organisation. As such, it is a thermodynamically open system which persists in far-from-equilibrium conditions, i.e. roughly, the system must constantly 'order' or 'constrain' the flow of energy and matter traversing it (E. Thompson 2007: 97f., Moreno and Mossio 2015: 5-9). Relatedly, what distinguishes the autopoietic system from its surroundings is not so much its physical boundary as the organisational identity which this boundary helps to maintain (Di Paolo 2009, E. Thompson 2011: 212ff.) (we return to this in Ch. 3). Indeed, there is an inherent 'primordial tension' (Di Paolo 2018) between the system's capacity for self-

production and self-*distinction*, respectively. Whereas the first, being reliant on surrounding matter and energy, requires the system to be open, the second demands it to be closed ('to keep it together') so as not to dissolve into its milieu (ibid.). In a cell, this tension is inscribed into its semi-permeable membrane, which selectively controls the flow of matter and energy and thus allows the system to be both open and closed.

According to MV, 'autopoiesis in the physical space is necessary and sufficient to characterize a system as a living system' (1973: 112). Given this statement, MV's view might be summarised as autopoiesis = life. More accurately, however, their position combines two claims: (a) 'autopoiesis is necessary and sufficient to characterize the *organisation* of living systems' (1973: 82. My emphasis), and (b) all other characteristic aspects of life, what MV call 'the phenomenology of living systems' (1973: 73, 84), are explanatorily dependent on the autopoietic organisation. Claim (b) can be further unwrapped in terms of MV's twofold rationale. It's because they think, first, that autonomy grounds the understanding of all other organic phenomena, e.g. reproduction or evolution, and as such is life's most essential feature, and, second, that autonomy can be captured in organisational terms, that they believe, on the basis of (a), that the phenomenology of life is grounded in autopoiesis. Thus, strictly speaking, MV do not think that autopoiesis is equivalent to the biological concept of life, but rather that it captures its essence or conceptual core.

1.1.3 Autonomy

On a closer look, however, matters are more complicated, due to the intricate relationship between autopoiesis and a further concept developed by Varela (1979), namely *Autonomy.* Unlike Varela, I am capitalising 'Autonomy' here – and will continue to do so – because Varela baptised this concept after the intuitive property it's supposed to explain, i.e. autonomy. **A**utonomy, however, is a technical concept, and not obviously equivalent to its namesake. Like autopoiesis, Autonomy articulates a particular kind of organisation of processes. As originally defined, a system is Autonomous iff its constituent processes

> (i) recursively depend on each other for their generation and their realization as a network,
> (ii) constitute the system as a unity in whatever domain they exist, and
> (iii) determine a domain of possible interactions with the environment
> (Varela 1979: 55, as quoted in Thompson 2007: 44)

Together conditions (i) – (ii) spell out the characteristic property of Autonomous systems: *organisational closure*.[3] In an organisationally closed system, every constituent process (a) enables and (b) is itself enabled by at least one other

[3] *Organisational* closure is sometimes distinguished from *operational* closure (Rudrauf et. al. 2003: 34; E. Thompson 2007: 45, Di Paolo et. al. 2017: 112). We can ignore this distinction and treat the notions as equivalent.

constituent process. Thus, an organisationally closed system is abstractly (i.e. formally) defined as a unity or individual entity consisting of all and only those processes which satisfy (a) and (b), constituting a self-sustaining network of processes which mutually support each other (Di Paolo 2009: 15f.). Condition (iii) (discussed later) is not, strictly speaking, part of the original definition of Autonomy by Varela (1979: 55). It is rather an implication of (i) and (ii) standardly cited in contemporary discussions, hence why, following Thompson (2007: 44), I include it here. Relatedly, although Autonomy has been extended to include the notion of precariousness (discussed later), Varela himself treated Autonomy and organisational closure as basically equivalent (Varela 1979: 55; Di Paolo and Thompson 2014: 71). Throughout, I will use 'Autonomy' strictly in the sense of (i) – (iii), as explained in this and ensuing sections.

Conditions (i) – (iii) are implied by, but more general than, conditions (i) and (ii) of autopoiesis. Hence, every autopoietic system is an organisationally closed network of processes and is, as such, Autonomous. However, not every Autonomous system is also autopoietic since there are many cases of organisationally closed systems which are not themselves autopoietic (Varela 1997: 81-87, E. Thompson 2007: 46, Di Paolo et. al. 2017, 2018). Thus autopoiesis is sufficient but not necessary for Autonomy.

To clarify this, we must examine what distinguishes autopoietic systems among the wider class of Autonomous systems. As Varela explains:

> The relations that characterize autopoiesis are relations of *productions* of components. Further, this idea of component production has, as its fundamental referent, chemical production […] it follows that the cases of autopoiesis we can actually exhibit […] have as a criterion of distinction a topological boundary, and the processes that define them occur in a physical-like space, actual or simulated in a computer. Thus the idea of autopoiesis is, by definition, restricted to relations of productions of some kind, and refers to topological boundaries. (1979: 54. Original emphasis).

Non-autopoietic Autonomous systems are not autopoietic because they lack a topological, spatially circumscribed *boundary*, e.g. a membrane, and their constituent processes are not processes of *production*, i.e. paradigmatically: chemical, molecular processes (cf. Varela, Maturana, and Uribe 1974). The closed network of interactions between members of an ant colony, for instance, establish the colony as an Autonomous system. But, on Varela's view, they are not processes of production nor do they generate a topological (physical) boundary.[4] Hence they do not qualify as autopoietic (Varela 1979: 54). The same exclusionary reasoning applies to other non-autopoietic Autonomous systems, e.g. nervous and immune systems (Varela 1979, 1997), sensorimotor networks (Barandiaran and Di Paolo 2014, Di Paolo et. al. 2017) and social and linguistic interactions (De Jaegher and Di Paolo 2007, Cuffari et. al. 2015, Di Paolo et. al. 2018).

What exactly distinguishes processes of *production* from other processes and how we should understand the notion of a topological boundary as opposed to a

[4] As E. Thompson notes, the colony's boundary is rather 'social and territorial' (2007: 44).

purely organisational boundary? We address these issues in Ch. 3. As a preliminary exercise, however, let us ask: are the distinguishing features of autopoiesis, i.e. (a) a topological boundary and (b) the biochemical, productive nature of the constituent processes, necessary to capture the *organisation* of living systems?

MV would say 'yes' given their contention that autopoiesis is necessary and sufficient to capture the organisation of living systems. However, given the concept of Autonomy, it seems less clear whether this is the right answer for the following reason: (a) and (b) do not seem to be *organisational* properties. Unlike Autonomy (organisational closure), they do not concern *relations* between processes. Rather, they specify processes of a particular *kind*, i.e. processes of production, and the form of their concrete manifestation, i.e. the construction of a topological boundary, as necessary conditions for autopoiesis. But one may question whether (a) and (b) contribute anything to the understanding of the autopoietic *organisation*. It seems one can ignore these features and still fully grasp how, say, the biochemical processes in a cell are organised, i.e. how they are recursively related. Indeed, given that autopoiesis is sufficient for Autonomy, all that is necessary to capture the autopoietic *organisation*, one might think, is the concept of Autonomy. For Autonomy spells out the relations of recursive dependence between the constituent processes of an autopoietic system irrespective of the nature and realisation of those processes (E. Thompson 2011: 215). What does the fact that these processes are molecular or chemical and form a concrete boundary contribute to our understanding of this organisation?

In response, MV might deny the claim that (a) and (b) are not organisational properties. After all, conceiving of organisation solely in terms of *relations* between processes seems quite restrictive. Can we not employ a broader notion of organisation according to which (a) and (b) do count as organisational properties? However, apart from the fact that it contradicts MV's own understanding of organisation (1980: xix, 76ff.) this strategy seems unsatisfactory for two reasons. First, it blurs what, intuitively at least, is an important theoretical distinction, i.e. that between an organisational *form* and its (material) *realisation*. Second, it doesn't answer the ultimately more important question: what do (a) and (b) contribute to the explanation of *autonomy* – life's self-sustaining capacity – which autopoiesis is supposed to deliver?

Alternatively, MV could concede that (a) and (b) are not organisational properties and hence that autopoiesis is unnecessary to capture the *organisation* of living systems, but maintain that autopoiesis is nonetheless needed to capture essential aspects of the autonomy of living organisms, which the formal concept of Autonomy doesn't capture. However, if (a) and (b) ground the distinction between autopoiesis and Autonomy, and if they are not organisational properties, this reply conflicts with an organisational approach towards life. It amounts to saying that there are non-organisational properties the understanding of which is necessary to capture what life is. This is consistent with MV's assumption that 'there is an organization that is common to all living systems' (1980: 76). But it is inconsistent with what they really want to say, namely that there is an organisation that is common to all *and only* living systems. But this thought is the core idea behind MV's organisational approach, namely that life can be fully captured in

organisational terms. The second strategy thus conflicts with the very spirit of MV's original proposal.

In raising these difficulties, I am not seeking to exegetically examine or reject MV's approach. I wish to highlight the first occurrence of a tension which runs deeply in the enactive approach. This tension arises from the ambiguity about the role which the notion of organisational form can play in our understanding of life's autonomy and how we should understand organisational form in relation to materiality. This tension will, in different guises, accompany us in what follows.

1.1.4 Structural Coupling

In an earlier paper, Maturana wrote:

> A cognitive system is a system whose organisation defines a domain of interactions in which it can act with relevance to the maintenance of itself, and the process of cognition is the actual (inductive) acting or behaving in this domain. *Living systems are cognitive systems, and living as a process is a process of cognition.* (1970: 13. Original emphasis).

To understand this striking passage, we must consider MV's conception of cognition which, in turn, is tied up with the notion of structural coupling. Earlier we saw how Autonomy formally defines a network of interdependent processes as an individual unity distinct from its surroundings. But we skipped an important aspect, encapsulated in condition (iii) of Autonomy, namely that, through forming an organisationally closed network of processes, an Autonomous system also defines a space of possible interactions with its environment. For instance, a cell will survive certain environmental encounters, e.g. uptake of nutrients, whereas others will kill it, e.g. high temperatures or consumption of poison. Calling such encounters 'perturbations', we can specify, more generally, a specific range of perturbations which are compatible with the cell's existence – what Di Paolo (2005: 438) calls the system's *viability set*. Condition (iii) of Autonomy expresses the thought that a system's viability set is determined by its organisation. For example, what counts as a nutrient or as poison, and hence what falls into or outside of the viability set, is relative to the cell's metabolism, which in turn depends on its organisation (i.e. the recursive dependencies between chemical processes). Some oceanic bacteria, for example, thrive on high concentrations of sulphur which are poisonous for other bacteria (Schulz et. Al. 1999). The status of *sulphur consumption* as a possible interaction in the bacteria's viability set is relative to the bacteria's organisation.

Once established as a unity, an Autonomous system is constantly exposed to perturbations. These perturbations naturally affect the system, they lead to 'structural changes' in MV's (1980: xx) words, e.g. sucrose consumption may alter the chemical composition or size of the bacterium. However, as long as these structural changes leave the system's organisation intact, its viability as an individual unity remains. *Structural coupling* refers to the history of interactions between one system and another, e.g. an organism and its environment, which

maintain a system's organisational identity over time (ibid.: xxi; Maturana 1975; E. Thompson 2007: 45). What is special about the structural coupling of autopoietic systems (and Autonomous systems more generally) is that, in them, any structural change following a perturbation serves to maintain autopoiesis. Which perturbations constitute the viability set and how the system can be changed by them without losing its Autonomy is fully determined by its organisation. In MV's words, there is a 'total subordination of the phenomenology of the system to the maintenance of its unity' (1973: 97).

Cognition, on MV's view, is the process by which an autopoietic system changes, in response to perturbations, so as to maintain its organisation. It is, in other words, the process which enables and shapes the structural coupling of a system with its environment. This is why MV say that '[t]he domain of all the interactions in which an autopoietic system can enter without loss of identity [i.e. its viability set] is its *cognitive domain*' (1973: 119. My emphasis). The interactions in this domain are, by definition, poised to maintain the systems organisational identity, and as such count as *cognitive processes* in MV's sense. Since, for MV, autopoietic systems are living systems, and living is grounded in the process of autopoiesis, Maturana's pronouncement follows: living systems are cognitive systems, and living as a process is a process of cognition. In other words, life is sufficient for cognition.

This is the first expression of what has since become known as *the deep continuity between life and mind*, a central idea of the enactive approach mentioned earlier (E. Thompson 2007: Ch. 6, Di Paolo 2018). Life-mind continuity has been understood in different ways, some of which we touch on here.[5] For now, it is enough to observe that MV's understanding of the life-mind relationship is based on a sufficiency claim: any living system is ipso facto a cognitive system. But is life also *necessary* for cognition? We return to this question later.

1.2 The Normative Turn

The new millennium brought what Barrett (2017) fittingly describes as 'the normative turn' in enactive theory. This development reaches back to Varela et. al.'s (1991) notion of *enaction* developed in *The Embodied Mind,* and receives its current form through the interconnected concepts of *sense-making*, *immanent purposiveness*, *adaptivity* and *precariousness*. We shall take each of these in turn.

1.2.1 Enaction

Whilst MV's autopoietic theory furnishes the conceptual cradle of enactive thought, it was not until *The Embodied Mind* (Varela et. al. 1991) that *enaction* as a novel approach towards cognition entered the cognitive sciences. Although autopoietic theory is left implicit in this book, it clearly informs Varela et. al.'s understanding

[5] See E. Thompson (2007: Ch. 6), Wheeler (2011), De Jesus (2016), Kee (2021), and Prokop (forthcoming) for discussion.

of cognition. A central tenet of their view is that 'cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided' (ibid.: 173).[6] This proposition is essentially a corollary of Varela et. al.'s view that cognitive systems are Autonomous (organisationally closed, self-organising networks) (ibid.: 150ff.). More specifically, we can understand it as an elaboration of condition (iii) of Autonomy, which, as explained, underwrites MV's idea of structural coupling. Varela et. al. follow MV's idea that cognition is a process of structural change of a system based on interactions with its environment, guided by the maintenance of its Autonomy. However, the notion of enaction emphasises the idea that the relevant structural changes are due to the *self-preserving activity* of the system itself. That is, cognition depends on what Varela et. al. call *embodied action* (ibid.: 173), i.e. action understood as the cause and product of the cognitive and sensorimotor structures of an Autonomous system. This subtle yet important point underlies Varela et. al.'s gloss of enaction as the 'bringing forth' of a cognitive domain (ibid.: 156; 165; 168), which has since become a sort of mantra for the enactive conception of cognition (see Sect. 1.2.2).

In *The Embodied Mind*, the role which the system's own activity plays in this 'bringing forth' or enactment of a cognitive domain and whether it can even be called an 'action' or 'enactment' is unclear. For MV, an Autonomous system's role in the structural coupling with its environment is *reactive*, it simply responds to perturbations so as to preserve its viability. The only principle guiding these responses is the maintenance of the system's organisation which defines its identity over time. Whilst these responses trigger processes which alter the system's structure or composition, e.g. as when bacteria change their metabolism in response to osmotic pressure, it is not *the system itself*, but its environment, which 'selects' the perturbations which the system will endure and which shape the process of structural coupling (i.e. cognition in MV's sense) (Barrett 2017). Thus, although there is a sense in which the system 'brings forth' its cognitive domain through its responses, the structures emerging from these interactions are ultimately subject to the environment's whim. In that sense, MV's Autonomous systems are not *agents* of their cognitive activity.

Varela et. al.'s (1991) conception of cognition seems to attribute more proactive capacities to Autonomous systems. However, it is not clear what warrants the author's change of perspective. For the theoretical framework on which the idea of cognition as enaction (embodied action) is based, involving the concepts of Autonomy, organisational closure and structural coupling, is not fundamentally different from MV's. Moreover, as Barrett (2017: 435) illustrates, Varela et. al.'s (1991: Ch. 9) discussion of development and evolution presents a system's self-preserving activity as following a 'proscriptive logic' according to which 'any action undertaken by the system is permitted as long as it does not violate the constraint of having to maintain the integrity of the system and/or its lineage' (ibid.: 205). Since the fundamental principle guiding the system's activity is to maintain its organisational identity, any structural change is permitted as long as this

[6] The other is that 'perception consists in perceptually guided action' (Varela et. al. 1991: 173). Here, we can ignore the topic of perception. See e.g. Noë (2004) for discussion.

organisation persists. But this again subjects the emergence of cognitive structures through structural coupling to ecological goings-on. In Barrett's words, 'while the system specifies the constraints of viability, the environment is what enforces those constraints […] [Varela et. al.] do not show how the proscriptive logic of viability constraints amounts to a normatively *self*-regulated form of structural coupling' (2017: 435. Original emphasis.)

More could be said, however, here it is sufficient to draw two general insights from our discussion. First, by reconceiving of cognition in terms of embodied action, Varela et. al. paint a picture of cognition as a process which depends on agency, i.e. a system's capacity to perform *actions*. Second, by thinking of a system's cognitive domain in terms of enacted constraints of viability, they present cognition as a process which, in Haugeland's words, 'abides in the meaningful' (1998: 230) i.e. as a process internally related to *norms* and *meaning*. In sum, *The Embodied Mind* leaves us with a conception of cognition tied to the concepts of agency, normativity and meaning, anticipating the current enactive notion of *sense-making*.

1.2.2 Kant, Sense-Making and Immanent Purposiveness

In an influential article, Weber and Varela (2002) endorsed an explicitly normative conception of organisms as self-concerned beings with their own ends and perspective, paving the way for the contemporary enactive conception of organisms as sense-making and immanently purposive systems. Their argument proceeds against the backdrop of Kant's views on organic teleology. So, let us begin with an all-too-brief summary of Kant's reflections in the third critique.

1.2.2.1 Kant on Natural Purposes

Kant famously argued that organised beings are *natural purposes/ends* [*Naturzwecke*]. According to Kant, 'a thing exists as a natural end **if it is cause and effect of itself** (although in a twofold sense)' (2000: 243. Original emphasis). Kant acknowledged a tension in this idea: how can something be *both* cause and effect of itself? To explain, he complemented his initial gloss with two further conditions.

> (1) 'its parts (as far as their existence and their form are concerned) are possible only through their relation to the whole.' (ibid.: 244f.).
> (2) 'its parts be combined into a whole by being reciprocally the cause and effect of their form.' (ibid.: 245).

If Beth builds a Lego dinosaur, that dinosaur only exists because Beth built it. On Kant's view, this presupposes that there is the *idea* of the *whole* dinosaur in some rational intellect, e.g. Beth's mind, which figures as the *final cause* of the existence of parts with a specific form arranged dinosaur-wise (ibid.: 245). This dependence on the idea of a whole distinguishes the Lego dinosaur as a thing which is only possible as an end.

However, the dinosaur, and artefacts more generally, are not *natural* purposes, i.e. cause and effect of themselves, since the cause of their existence is extrinsic. It lies outside themselves in the reason and concepts of an intelligent being (ibid.). But how could a thing be possible only as an end, by being brought under the idea of a unity, without thereby being caused from outside, and hence be an artefact? Here, (2) takes hold: to be a natural purpose, a thing's parts must not only depend on the idea of the whole, they must also mutually cause and effect each other. As Kant explains,

> [f]or in this way alone is it possible in turn for the idea of the whole conversely (reciprocally) to determine the form and combination of all the parts: not as cause – for then it would be a product of art – but as the ground for the cognition of the systematic unity of the form and the combination of all the manifold that is contained in the given material for someone who judges [*beurtheilt*] it. (ibid.).

Only if the parts of the whole can be seen as being cause and effect of each other, by mutually reproducing themselves, do they depend on the idea of a whole *without that idea being the cause for their existence*. Only then can the thing be seen as both cause and effect of *itself*, as a 'self-organized being' (ibid.), i.e. a natural purpose.

But how is this possible, if – given (1) – the parts still depend for their existence on the idea of the whole? Kant's answer is interesting: yes, they do depend on that idea, however, not *causally* – which would make the thing an artefact – but rather *cognitively*. That is, the idea of the whole is necessary in that someone *judging* the whole from outside must possess that idea in order to grasp the self-organizing manifold as a unity, i.e. an organism (E. Thompson 2007: 134f.). In this way, Kant hoped to resolve the puzzle of characterising something as both cause and effect of itself. It is cause in that the parts *effectively (mechanically) cause* each other, and thereby bring about the whole – this is what makes the thing *natural*. It is effect in that the idea of the whole appears as the *final cause* of the organized unity of the parts for someone judging it – this is what makes the thing a *purpose*. In a natural purpose, the two kinds of causality – *nexus effectivus* and *nexus finalis* – combine into a whole, e.g. an organism, which is both cause and effect of itself 'in a twofold sense'.

Kant's analysis connects to two further points. On the one hand, although the organism appears as an end in itself only to someone judging it, the purposiveness is still *intrinsic* in that we, as observers, are forced – by a 'remote analogy' (2000: 247) to our own final causality – to behold the organism as a purposeful being in order to grasp it as a unified, self-organizing whole. On the other hand, this intrinsic purposiveness lies still somewhat outside the 'true' nature of organisms. This is because Kant's view of nature was mechanistic (following Newton), and he deemed it impossible to explain the teleological structure of organisms, i.e. natural purposes, mechanistically (ibid.: 270f.). In sum, Kant articulates the idea of organisms as natural purposes: self-producing, self-organizing beings, which must be understood as intrinsically purposeful wholes. However, Kant also presents a conception of organic teleology as a necessary mode to apprehend organisms, rather

than a feature which could be explained naturalistically and thereby established to be at home in nature, rather than outside it.

1.2.2.2 Sense-Making and Immanent Purposiveness

Drawing on Jonas' philosophical biology, Weber and Varela espoused a conception of organic teleology which exceeds Kant's. Jonas regarded the teleology of organisms as 'not just an alternative choice of description' (2001: 91), but as an inherent feature of life. For Jonas, organisms are essentially *metabolising* entities. As such, their ongoing existence depends on the constant throughput of matter from their surroundings. Consequently, Jonas saw organisms as fundamentally *concerned* individuals, which purposefully strive to uphold their living form through their own, self-preserving activity (2001: 79-86). We expand on Jonas' analysis in Ch. 3.

Following Jonas, Weber and Varela (2002) argued that living organisms are fundamentally self-concerned, purposeful beings with a meaningful perspective on the world. They also claimed that this Jonasian conception of organic teleology receives a naturalistic foundation in autopoietic theory. Their position builds on earlier work by Varela (1991, 1997), who proposed two interrelated ideas as essential to the view of organisms as autopoietic, Autonomous systems:

> (1) Organisms are fundamentally a process of constitution of an identity.
> (2) The organism's emergent identity gives, locally and mechanistically, the point of reference for a domain of interactions. (1997: 73)

These statements are meant to complement each other, 'they are supposed to describe two sides of one emergent process' (E. Thompson 2007: 147). Let us illustrate this using Varela's example of bacteria swimming up a sucrose gradient. When placed in a sucrose solution, certain bacteria swim towards the area of highest sugar concentration by beating their flagella. A natural way to explain this behaviour is to understand sucrose as an attractor, e.g. a *nutrient* for the bacterium (required to uphold its metabolism). However, in adopting this explanation, we have implicitly adopted the bacterium's perspective. Nothing about the chemical composition of sucrose itself links the molecule to the bacterium. In Varela's terms, sucrose only counts as a nutrient, and hence as valuable, by reference to the bacteria's Autonomous identity (1997: 79). More generally, the domain of interactions, e.g. approach, escape, or contraction, in (2) is determined in relation to the organism's identity whilst equally enabling the metabolic process of identity-constitution in (1), which again shapes the organism's interactions with its environment. Thus, for Varela, the concepts of identity and interaction form a co-dependent pair, the yin and yang characterising the dynamic mode of being of living things.

Varela's view elaborates the normativity characterising the relation between organism and environment, implicit in the idea of enaction in *The Embodied Mind*. Through the constitution of an identity, the environment becomes a 'world', i.e. (roughly) the totality of existentially relevant structures and possibilities of

interaction which are meaningful in relation to that identity (compare the related notions of a system's 'cognitive domain' or 'viability set'). In Varela's enigmatic words '[t]he difference between environment and world is the surplus of signification which haunts the understanding of the living and of cognition, and which is at the root of how the self becomes one […]'. (1997: 79).

Varela (1991, 1997) more explicitly emphasises the system's *active* role in constituting its world. In constraining the flow of molecules across its membrane or moving towards sucrose, the bacterium is not merely following a genetically pre-determined programme or responding to perturbations. Rather, it is actively endowing sucrose with significance *through* its activity, e.g. by approaching it. In this sense, the bacterium creates its own world, it enacts and shapes a meaningful environment through its own activity. This is the core thought behind the contemporary enactive understanding of basic cognition as *sense-making*, i.e. an individual's bringing forth of a cognitive domain (a world) on the basis of its Autonomy (E. Thompson 2007, Thompson and Stapleton 2009, Di Paolo 2009, Di Paolo and Thompson 2014).

Varela (1997: 80) initially denied that sense-making implies any teleology in an Autonomous system (cf. Maturana and Varela 1973: Ch. II). However, following Jonas, Varela (and Weber) eventually tied sense-making to a conception of organic teleology which goes 'beyond Kant' (Weber and Varela 2002: 212) in three respects. First, it is realist: organisms are claimed to have their own, intrinsic purposes irrespective of an observer, grounded in their fundamental concern to affirm their own identity (following E. Thompson (2007: 146, 2011: 211), I will call this teleology *immanent purposiveness/teleology* to distinguish it from Kant's conception). Second, it is subjective: by self-producing and being concerned with their own identity, organisms embody a perspective which is their own. Finally, it is normative: organisms stand in a meaningful relation to their environment, i.e. their world, which they imbue with significance on the basis of their self-concerned identity and purposeful activity (Weber and Varela 2002: 109-121). With this in mind, *sense-making* may be more precisely characterised as the activity of a system which is (a) *immanently purposeful*, i.e. aiming to maintain and reproduce its Autonomous identity (b) *subjective*, i.e. embodying a self-concerned perspective and (c) *normative*, i.e. bringing forth a meaningful, cognitive domain (Thompson and Stapleton 2009, Di Paolo 2009, Di Paolo and Thompson 2014).

1.2.3 Adaptivity

Weber and Varela's 'immodest conclusion' (2002: 120) that living organisms are immanently purposeful, sense-making beings spawned much controversy.[7] Arguably the most important developments are Di Paolo's proposal of the concepts of *adaptivity* (2005) and *precariousness* (2009). Let us begin with the first.

[7] Critics mainly object to the realist and subjective elements of the Jonasian, enactive conception of organic teleology, corresponding to (a) and (b) of sense-making. See Ward and Villalobos (2015, 2016a, 2016b), De Jesus (2016), Kee (2021), and Di Paolo et. al. (2018: 33-36) for discussion.

Like Barrett's (2017) criticism regarding *The Embodied Mind*, Di Paolo argues that autopoietic theory, particularly the concepts of Autonomy (organisational closure) and structural coupling, do not justify Weber and Varela's claims about the kind of self-preserving activity which sense-making implies. Sense-making presupposes that an autopoietic system can engage in forward-looking, pre-emptive activity in order to maintain its organisational identity, e.g. swimming up a glucose gradient, to improve its viability conditions. However, as Di Paolo argues, we cannot explain this 'projective teleology' (2005: 433) through autopoiesis alone, as Weber and Varela suggest. For autopoietic systems are, in Di Paolo's terms, *robust* but not *adaptive* (2005: 434-437). That is, they can endure a certain range of perturbations (those in their viability set) without losing their organisation. But autopoiesis is a concept without degrees, i.e. either a system satisfies the autopoietic organisation or it doesn't. Consequently, from the viewpoint of the autopoietic system we cannot distinguish between different kinds of perturbations, apart from those which destroy its organisation and those which don't. Metaphorically speaking, an autopoietic system's world consists of life or death and nothing in between. There is no 'reason' for the system to engage in pre-emptive, viability-conducive activity as long as its organisation is conserved, i.e. as long as it's *alive* in the only sense that autopoiesis provides (ibid.).

To make sense of sense-making, Di Paolo complemented autopoiesis with the notion of *adaptivity*. In short, an adaptive system has the capacity to (a) identify tendencies (in itself or the environment) as moving it closer to or away from its disintegration, i.e. the breakdown of its organisation and (b) to regulate its activity based on the identification of these tendencies so as to improve its conditions of viability (2005: 438). Adaptivity combines two capacities, corresponding to (a) and (b). The first is the *evaluative access* to internal and external events in relation to its boundary of viability. This is essentially a discriminatory capacity, enabling the system to distinguish a variety of features according to their (potential) existential relevance for it. The second is the *practical ability* to regulate its own activity in a reactive and, more importantly, *pro*active fashion in relation to these features so as to preserve its organisation. This is at bottom an agentive capacity, enabling the system to engage in the forward-looking kind of activity which sense-making implies.

In Di Paolo's view, these two capacities are internally related. That is, evaluative discriminations are without content unless they are linked to adequate regulative responses, and projective activity is blind without the identification of existentially relevant goings-on (2005: 438f.). In a further move, Di Paolo argues that the normative, evaluative dimension characterising the relation between a system and its world constitutively depends on its self-preserving activity as characterised by this twofold adaptive capacity: '[b]oth elements, self-monitoring and appropriate regulation, are necessary to be able to speak of meaning from the perspective of the organism.' (2005: 438).

This conceptual marriage between normativity and adaptive, purposeful activity underlies an interesting aspect of Di Paolo's dialectic, namely that a general conceptual point about *our understanding* of the normative dimension of sense-making is used to motivate the introduction of a *capacity* which, Di Paolo argues,

illuminates the possibility and constitution of this normative domain on the basis of the organism's own activity. That is, adaptivity is primarily introduced to explain how, from the organism's perspective, graded normative evaluations of its current state of viability are possible. For the possibility of such evaluations is the *precondition* for the organism's engagement in any forward-looking, teleological activity. However, Di Paolo wants adaptivity to do more than that, namely to also account for the *grounding* and *construction* (sense-*making*) of a graded normative domain which autopoiesis cannot provide, but which sense-making requires. It's precisely because, on the enactive view, a normative domain is brought forth by and thus depends on an organism's adaptive, purposeful activity that Di Paolo suggests that adaptivity accounts (together with autopoiesis) for the grounding and generation of this domain. In Ch. 3, we examine whether, given a certain conception of the identity of Autonomous systems, Di Paolo's proposal lives up to its promises.

1.2.4 Precariousness

Are snowflakes Autonomous? The answer is not as obvious as it appears. Given certain weather conditions, snowflakes form unique, wonderfully elaborate structures and one might argue that these structures are organisationally closed and hence Autonomous. If we think of the flake's constituent molecules as 'processes', each might be said to depend on at least one other molecule for its own position whilst also enabling the position of some other(s). Thus, together the constituents form a network distinguishable from its surroundings through its closed organisation.

The concept of precariousness excludes such odd cases of Autonomous systems by adding the following condition: 'In the absence of the enabling relations established by the operationally closed network, a process belonging to the network will stop or run down.' (Di Paolo and Thompson 2014: 72). In a precarious system, the individual constituent processes cease to be if removed from the system. Precarious constituent processes are like infatuated lovers: they can only exist together and perish in isolation.[8] A snowflake is not precarious because, although its structure resides in the co-dependent relations between its constituent molecules, the molecules do not themselves rely on these relations to exist. They will persist such as they are after the snowflake dissolves.

Precariousness also performs a further job which is, in some ways, more important than weeding out trivial cases of organisational closure (Di Paolo et. al. 2017: 116). Recall that adaptivity was introduced to explain the adaptive, purposive activity implied in sense-making. In that, it is an *addition* to the concept of autopoiesis, not a consequence of it (Di Paolo 2005: 438). But if autopoiesis and adaptivity are both central to understanding life and sense-making, these two concepts should be internally related, rather than separate additions. Precariousness illuminates this connection, i.e. it explains why the adaptive, purposive activity of sense-making is an implication, not a contingent feature, of the mode of being of

[8] Thus precarious systems misunderstand genuine love in Fromm's (1962) sense.

autopoetic systems.[9] Since the constituent processes of a precarious system tend to collapse when isolated from the network, the system is forced to actively uphold its organisation. It must constantly counteract the forces of decay acting on its constituents by maintaining the recursive organisation on which their persistence depends. A precarious, organisationally closed system is 'literally self-enabling' (Di Paolo and Thompson 2014: 72): it actively creates the conditions of its own existence (Di Paolo 2009: 16). Since, for this, it requires energy and matter from its surroundings and the ability to screen and regulate its current state of viability, this need for active maintenance requires it to adaptively relate to its environment.

Whilst elucidating the relation between autopoiesis and adaptivity, the standing of precariousness itself in relation to these concepts as well as Autonomy is ambiguous. Sometimes, precariousness is introduced as a *necessary* condition, besides organisational closure, for Autonomy (Di Paolo 2009: 15, Thompson and Di Paolo 2014: 69ff.). Sometimes, it is presented as a *clarification* of Autonomy, to exclude trivial cases of closure and explain why autopoiesis, and Autonomy more generally, must be adaptively maintained (Di Paolo et. al 2017: 116, Di Paolo et. al. 2018: 25f.). Furthermore, on the enactive view, precariousness seems to be a necessary condition for *life* (Di Paolo 2009: Fn. 5, Froese 2017, Di Paolo 2018), but its relation to autopoiesis is, again, somewhat unclear. I will continue to use Autonomy in the sense of conditions (i) – (iii), explained earlier, and thus as conceptually independent of precariousness. This does not mean that these concepts are unrelated. On the contrary, it enables a non-circular discussion of their relationship, which will become important in Ch. 3.

1.3 Agency

The concept of agency, left implicit thus far, occupies a key position in the enactive conceptual landscape. Sense-making is particularly closely related to agency, as we shall see. The enactive approach to agency, which will play a central role in what follows, proceeds in two steps.[10] The first is the articulation of three necessary and jointly sufficient conditions specifying a general conception of agency. The second is what Barandiaran et. al. call a 'generative definition' (2009: 376), i.e. a formulation of the requirements that a system must meet to be an agent, which explains, in naturalistic terms, how the features of the general conception emerge from the dynamical properties of the system.

[9] Thanks to Ezequiel Di Paolo for helpful discussion on this point.

[10] 'The enactive approach to agency' here means the approach developed in Barandiaran et. al. (2009) and Di Paolo et. al. (2017). This conception owes much to Barandiaran (2008). However, I will mostly refer to Di Paolo et. al. (2017) since it contains the most comprehensive version of the approach.

1.3.1 Trisecting Agency

The general conception of agency consists of a trio of requirements: self-individuation, interactional asymmetry, and normativity (Barandiaran 2008: 82-88, Barandiaran et. al. 2009, Di Paolo et. al. 2017: 111-127). Let's consider these individually.

1.3.1.1 Self-Individuation

It is widely accepted that, as Aristotle suggested, an action is a kind of *self*-movement (see e.g. Coope 2007, Hornsby 2004). But this raises the question of what (if anything) 'self' refers to. In contemporary action theory (elaborated in Ch. 2), the question of what distinguishes an agent as an individual capable of performing actions – the subject S in the schema 'S is doing A' – is rarely explicitly addressed, at least not beyond the topic of personal identity. Yet, even the best theory of personal identity is unlikely to provide a generally acceptable account of the individuality of agents. For there are many beings, e.g. new-borns, birds, cats and perhaps certain artificial systems like *AlphaGo*, which can, intuitively, *do* things despite falling short, to different degrees, of personhood. So what, generally speaking, makes an agent an individual?

Simply put, the condition of self-individuation requires that, for a system to be an agent, there must be a clear answer to this question. Specifically, the answer must, in some sense, be given by the agent itself (Di Paolo et. al. 2017: 111ff.). Its individuality cannot exist only insofar as an observer *thinks* of the agent as an individual. Rather, to be a genuine individual, an agent must *self*-individuate, i.e. define its identity by itself, independently of an external gaze (Jonas 1968, 2001: 79). To illustrate this, consider a distinction which underwrites our intuitive understanding of agency, namely that between what the agent does and what someone thinks the agent does.[11] When a court hears someone's testimony about the actions of the defendant accused of murder, they want to determine whether what *the defendant* did amounts to murder. In that, they presuppose a distinction between the defendant's actions and what others think about his actions. Without this distinction, the question whether *he* did it would be unintelligible. But to understand this distinction, it seems, we need an account of what individuates the defendant as an agent and does so irrespective of someone else's conceptions about him. The self-individuation requirement registers this point. It demands that an agent be individual which actively and independently distinguishes itself from its environment (Barandiaran et. al. 2009: 369f.; Di Paolo et. al. 2017: 111-116).

[11] Davidson presupposes this distinction when writing that '[a] reason rationalizes [explains] an action only if it leads us to see something the agent saw, or thought he saw, in his action' (2001: 3). A reason explains an action not by showing how it makes sense from the perspective of an onlooker, but by appealing to *the agent's perspective*. This presupposes that we understand whose perspective we are talking about.

1.3.1.2 Interactional Asymmetry

When explaining what an action is, or why someone did something, we tend to look towards the agent, not its surroundings. Why? One central reason is the assumption that agents are the source or cause of their actions and, as such, hold the key to explaining why a happening qualifies as an action. This assumption is so central that, when a theory explains actions through causal relations which, upon reflection, do not involve *the agent* herself, we are inclined to deny that it is a theory of *agency* at all (Horsnby 2004).

The interactional asymmetry condition accommodates this intuition by demanding that, for a system to be an agent, it must be capable of being the source of its interactions (with the environment). Interactional asymmetry contrasts with what we might call interactional *symmetry*. MV's notion of structural coupling, for instance, is symmetrical because it describes the co-dependence between a system and its environment without designating either as playing a privileged, leading role in bringing this structural dependence about (Di Paolo et. al. 2017: 116.). Regarding the origination of their interactions, the system and its environment are on a par.[12] In order to speak of interactional asymmetry, then, there must be something about the system which justifies taking it as the *source* of its interactions (Barandiaran et. al. 2009: 270-272; Di Paolo et. al. 2017: 116-120).

1.3.1.3 Normativity

Many winters ago my friends and I built a daring snow ramp. I gained speed, took off, and woke up minutes later with a broken collarbone. What I *wanted* or *intended* or *tried* to do was to gracefully land on my snowboard. That was the standard or goal my action hoped to live up to. My becoming hurt was, instead, an accident, something that happened, not something I *did*, i.e. an *action*. This possibility of an action's failing (or succeeding) to meet its standard or goal is essential to actions. An action is *as such* something which can succeed or fail (in different ways – we return to this in Ch. 2). Put differently, actions are always subject to certain *norms* which govern their performance, and relative to which they can be evaluated.

Roughly, norms can apply to an agent in two ways (Rödl 2007: 20). They can be *internal*, i.e. they can apply to the agent in virtue of the kind of thing it is. When we say, dismissively, 'this pianist keeps playing the wrong notes' we are evaluating the agent *as a pianist*. Insofar as she *is* a pianist, our evaluation applies to her, e.g. she cannot encounter 'Oh, but who said I should play the right notes?'. It is of the nature of a pianist to play the right notes. Hence, insofar as she is a pianist, the norm *play the right notes* applies to her. Conversely, if the agent was a dog, say, the judgement would be based on *external* norm. Since a dog is not a pianist, its keyboard-stomping cannot be evaluated as *playing the wrong notes* based on a norm which is internal to what it is to be a dog.

[12] If Barrett (2017) is right, then it is the environment, if anything, which plays the causal lead in structural coupling. See Sect. 1.2.1.

The normativity requirement demands that, for a system to be an agent, its activities must be subject to norms or goals which are its own, i.e. which are internal or 'intrinsic' in Di Paolo et. al.'s (2017: 121) terms. Put differently, the normativity of agency must be independent of the evaluative judgements of an observer. It must be internal to the agent's own nature (ibid.: 120). Normativity, thus understood, parallels Foot's (2001) notion of *natural goodness*, discussed later.

1.3.2 Grounding Agency

Given these three requirements, Di Paolo et. al. (2017: 127) propose the following generative definition of agency (following Barandiaran et. al. 2009: 376). A system *S* is an agent engaged in a coupling *C* with an environment *E* if and only if:

1. *S* is an *autonomous* system, meaning that:
    a. *S* is an operationally closed network of precarious processes whereby every process belonging to the network is enabled by at least another process of the network and enables at least one other process in it, so that isolated from the network any component process would tend to run down or extinguish;
    b. *S* actively and functionally distinguishes itself as a unity and the set of processes (not belonging to *S*) that can affect *S* and are affected by *S* defines *S*'s environment (*E*); and
2. *S* sometimes exercises a capacity to *modulate* the coupling C in an *adaptive* manner:
    a. where *modulation* indicates an alteration (dependent on the state of *S*) in the set of parameters and conditions that affect the coupling between *S* and *E*;
    b. and *adaptive* means that modulations in the coupling *C* contribute to keeping *S* as a viable system. (Di Paolo et. al. 2017: 127. Original emphasis).

(1) spells out the self-individuation requirement in terms of Autonomy and precariousness. That is, an agent self-individuates through being an organisationally closed, precarious system (1a). The mention of 'functional' self-distinction in (1b) demands, additionally, that some of the system's properties are dedicated 'to keeping parts of the system partially isolated from the environment in the context of organizational requirements' (Di Paolo et. al. 2017: 115). The authors do not say much about the idea of *functional* self-distinction (in contrast with the *formal* notion of distinction provided by organisational closure) but the idea seems to be that, for self-individuation to obtain properly, it is insufficient that a system merely maintains its organisation. Its self-individuation must also be concretely expressed in specific mechanisms or structures which serve to create 'a minimal spatiality, a difference between an inside and an outside space' (ibid.), e.g. a semipermeable boundary. This addition of functional self-distinction is interesting, since, first, it is absent from Varela's definition of Autonomy and, second, it comes suspiciously close to Varela's notion of a topological boundary, which is one of the criteria distinguishing autopoiesis from Autonomy. We return to this later.

Condition (2a) addresses the requirement of interactional asymmetry. To see how, we must understand the phrase 'modulation of the coupling C'. As explained, for interactional asymmetry, there must be a feature of the system which justifies our taking it as the *source* of its interactions. As Di Paolo et. al. (2017: 117f.) note, one might seek to capture this feature in energetic or statistical terms by appealing, for example, to a system's ability to curtail the flow of energy across its boundary or to probabilistic, time-dependent markers indicating an increase in systemic activity relative to goings-on in its environment. According to the authors (ibid.: 118), such approaches aim to introduce an asymmetry *within* the system-environment relation by specifying peculiarities of one relatum, i.e. the system. The authors instead conceptualise interactional asymmetry as a system's capacity to affect *the system-environment relation itself*, i.e. to directly influence the structural co-dependencies between systemic and environmental processes. By the system's *modulation* of its coupling, they mean the system's alteration of the coupling relation itself, not of its environment directly (ibid.: 117ff.).

Consider the following example. During drought, resurrection plants detach their roots and turn into a shrivelled ball light enough to be carried by desert winds. From an energetic or statistical standpoint, the curled-up plant cannot be regarded as the originator of its motion since its movement entirely depends on the climate-dependent energy in the air. However, that the air is able to carry the plant at all is the result of the plant's alteration of its environmental coupling, and it is in this sense that, on the enactive view, the plant is responsible for its movement. The plant's curling-up introduces an asymmetry into the plant-environment relation not because of its energetic or statistical profile, but because it directly affects *the structure of this relation itself*, i.e. 'the set of parameters and conditions' in (2a), by significantly altering the way in which wind can affect the plant.

This example also illustrates that the applicability of the criterion of interactional asymmetry and of the enactive conception of agency more generally is context-sensitive relative to the level and time of observation (Di Paolo et. al. 2017: 117-120; 128). A desert traveller ambushed by a flying resurrection plant will be unable to discern the plant's active role in this interaction because her period of observation excludes the plant's alteration of its environmental coupling. Similarly, a biochemist analysing the plant's metabolism might overlook this asymmetrical alteration because the processes which enable the plant enfold do not show up at that level. This context-sensitivity does not imply that interactional asymmetry or agency are not real, observer-independent phenomena. As Di Paolo et. al. (ibid.) stress, once the time and scale of observation is set, there will be a clear answer regarding whether the observed events add up to an asymmetrical interaction or not. It might of course be difficult to determine the right level and duration of observation – and here statistical and energetic signatures might help (Di Paolo et. al. 2017: Fn. 17) – but this doesn't mean that the phenomenon itself is observation-dependent. You just have to look at it in the right way to see it (cf. Dennett 1991).

Condition (2b) concerns the requirement of normativity. An *adaptive* modulation, which Di Paolo et. al. (2017: 120) call a *regulation*, of the coupling is distinguished from a mere modulation by contributing to S's viability. A regulation is *normative* by being a modulation carried out in view of keeping the system's

viability intact. Thus, unlike mere modulations, *regulations* can fail or succeed; they can be *evaluated* relative to how they alter the system's position with respect to its boundary of viability. The resurrection plant's curling-up, for example, is a regulatory response to drought because it contributes to the plant's survival. If the plant does not furl quickly enough and dies as a consequence, its regulatory response can be said to have *failed*.

To fully unpack the idea of normativity in (2b), we need to understand how a system's modulations relate to its viability and what 'viable' means. Luckily, we have already discussed this in relation to sense-making. That is, organisational closure provides a basic norm in that the system must actively maintain it in order to exist. Roughly, 'viable' means 'organisational closure is intact' and the system's modulations can be evaluated relative to whether or not they contribute to upholding closure. However, as discussed earlier, adaptivity is added to this all-or-nothing norm to explain how graded evaluations and adequate, future-oriented regulations are possible from the system's perspective (Di Paolo et. al. 2017: 120-124). Thus, 'the capacity to *modulate* the coupling C in an *adaptive* manner' in (2) encapsulates the enactive conception of normativity understood as the evaluative dimension characterising a system's cognitive domain or 'world' which the system itself brings forth through its activity. We return to normativity in Ch. 3.

Finally, as Di Paolo et. al. (2017: 119f.; 123) emphasise, to qualify as an agent, an Autonomous system need not engage in adaptive modulations non-stop, nor is it necessary that its regulations always succeed in improving its viability. As (2) specifies, the system must only have the *capacity* to regulate its interactions and 'sometimes' realise it.

1.4 Summary and Open Questions

After this *tour d'énactivisme*, let us review the enactive concepts and their relations. It seems clear that cognition (sense-making) is *dependent* on agency, i.e. a system's capacity to perform actions. Indeed, given that, standardly, Autonomy (including precariousness) and adaptivity are considered to be individually necessary and sufficient for sense-making (E. Thompson 2011: 211, Thompson and Stapleton 2009: 25), one might wonder what distinguishes agency from sense-making. Although the emphasis on interactional asymmetry and regulation is not explicit in sense-making, both agency and sense-making are grounded in the same concepts, i.e. Autonomy, adaptivity and precariousness.

Here, it is sufficient to observe that agency is *necessary* for sense-making. For surely the capacity for 'behaviour and conduct in relation to norms of interaction which the system itself brings forth on the basis of its adaptive autonomy' (Di Paolo and Thompson 2014: 73), i.e. sense-making, presupposes that the system is capable of 'adaptively regulating its coupling with the environment according to norms established by its own viability conditions' (Di Paolo et al. 2017: 127), i.e. agency. Indeed, the dependence of sense-making (cognition) on self-preserving, purposive interactions is arguably the key insight behind the idea of *enaction* from which enactivism derives its name.

Another important point is that, since autopoiesis is sufficient but not necessary for Autonomy, autopoiesis is *prima facie* not necessary for sense-making. Since, generally, life is defined as adaptive autopoiesis, i.e. autopoiesis and adaptivity are individually necessary and jointly sufficient for life, this means that *prima facie* life is sufficient but not necessary for sense-making (adaptive Autonomy) (E. Thompson 2011: 211). Strictly speaking, though, this only holds if *adaptivity* likewise does not necessarily depend on autopoiesis. If it does, then, since adaptivity is necessary for sense-making, adaptive autopoeisis (life) would, after all, be required for a system to be a sense-maker.

Parallel points apply to the relationship between life and agency. According to Barandiaran et. al.

> Minimal life forms already come to satisfy the necessary and sufficient conditions for agency. This does not imply, however, that living organization is necessary for agency, nor that all forms of agency need to trace their normative or individuality conditions back to living organization. (2009: 376).

This claim, i.e. that the living organisation is sufficient but not necessary for agency, relies on the assumption that adaptive Autonomous systems are conceivable which satisfy the conditions for agency – individuality, interactional asymmetry, and normativity – independently of any relation to *living* systems or processes. Since, again, life is adaptive autopoiesis, this means that, for this claim to be true, systems must be conceivable which satisfy the conditions of agency *independently of autopoiesis* (and it must also be true that adaptivity does not require autopoiesis). On the enactive approach presented thus far, there is conceptual space for this possibility given that autopoiesis is *prima facie* not necessary for adaptive Autonomy.

Now, all the 'prima facies' of the preceding paragraphs not only sound philosophically sophisticated, they indicate some loose ends in the enactive web of concepts (which I will attempt to tie up in Ch. 3). To pluck these properly, let us highlight the following key enactive claims (see **Figure 1** below) (cf. E. Thompson 2011: 211f.; 2007: 158).

(1) Autopoiesis and adaptivity are individually necessary and jointly sufficient for life.
(2) Autonomy and adaptivity are individually necessary and jointly sufficient for sense-making (basic cognition).
(3) Autopoiesis is sufficient for Autonomy.
(4) Agency is necessary for sense-making.
(5) Life (adaptive autopoiesis) is sufficient for agency.
(6) By (1), (2) and (3), life is sufficient for sense-making (basic cognition).

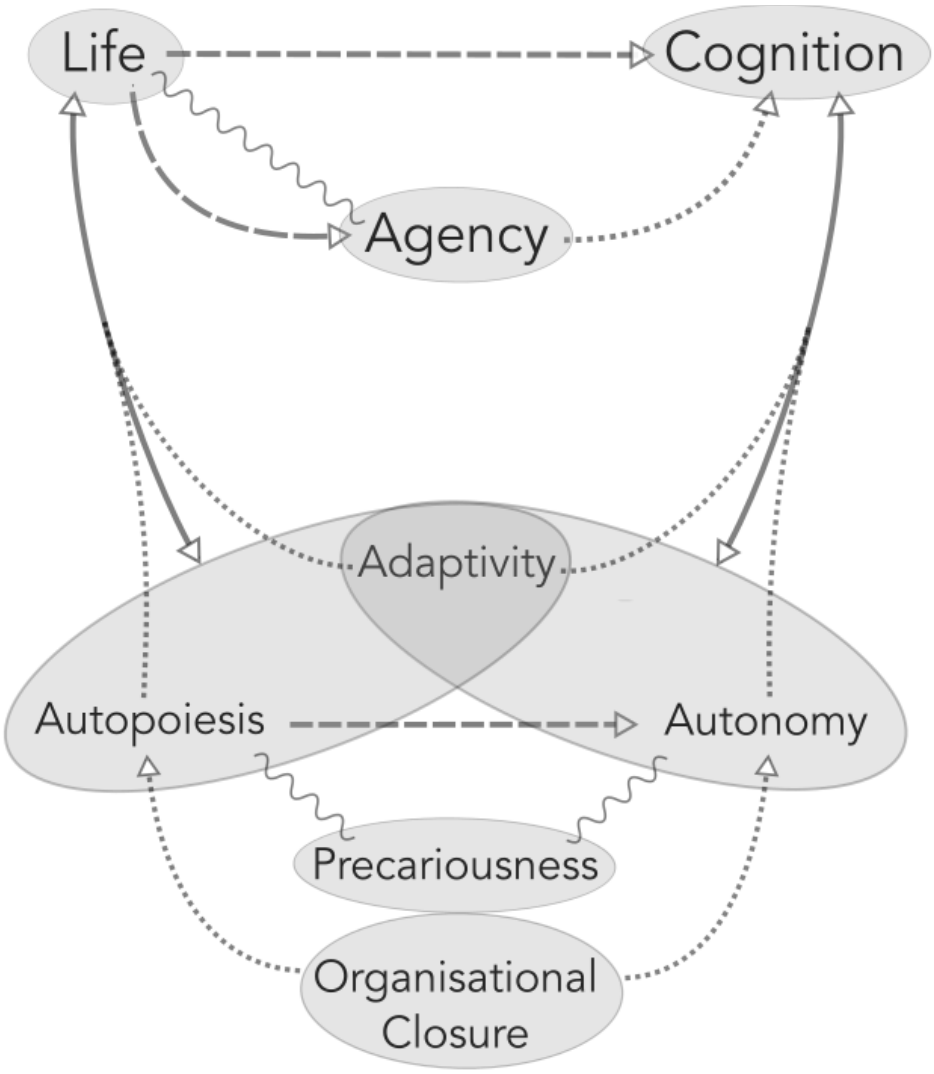


Figure 1 Conceptual map of the enactive approach. Conceptual relations are represented as follows: dotted lines indicate *necessity*, dashed lines *sufficiency*, bold lines *necessity and sufficiency* (equivalence), and wavy lines *uncertainty* (to be determined).

Taken together, these claims entail that life is *sufficient* for both agency and sense-making, yielding another version of Maturana's (1970: 13) claim that the process of living coincides with cognition. It is, in other words, one way to understand the continuity between life and mind, and – we may add – between life and action (cf. E. Thompson 2011, 2007: Ch. 6). However, (1) – (6) leave two important questions open:

(A) Is life necessary for agency?
(B) Is life necessary for sense-making (cognition)?

Prima facie enactivism has the conceptual resources to answer both questions negatively. What enables this denial is the relationship between autopoiesis and Autonomy. It's only because, conceptually speaking, autopoiesis is not necessary for Autonomy (or adaptivity) that there is logical space for saying that there might be sense-making (i.e. adaptive Autonomous) systems whose agentive, purposeful activity is not dependent on the living (i.e. adaptive autopoietic) organisation.

In the following chapters, I hope to narrow this conceptual space by arguing that, *insofar as the Autonomy of agents is concerned*, Autonomy depends on adaptive autopoiesis. To this end, the next chapter develops a general conception of agency, or more specifically of what I will call a *productive process* or *system*, following Anscombe's (2000) approach in *Intention*. The resulting conception will allow us to articulate two constraints which any theory of agency should meet. In Ch. 3, I then argue that, to meet these constraints, the enactive conception of agency must regard the adaptive Autonomy of agents as inextricably linked to adaptive autopoiesis. In that sense, life will be necessary for agency – and also, given (4), for sense-making, although this claim will remain mostly implicit. Let us, then, move on to Anscombian ground.

2 Productive Being

2.1 Why Anscombe?

In contemporary analytic philosophy of action, there is a rough division between the 'standard approach' to agency and alternative views, including recent interpretations of Anscombe's work (Ford. et. al. 2011, M. Thompson 2008). Donald Davidson, himself greatly influenced by Anscombe (Stoutland 2011), is the main exponent of the standard approach (Davidson 2001). The differences to Anscombe are significant, notably regarding her and Davidson's conceptions of causation and reason explanation (Hornsby 2011, Schlosser 2019). Here, the most relevant difference is the explanatory set-up according to which Anscombe and the standard approach hope to illuminate the notion of action. Both assume that the notion of action derives from and in this sense is less fundamental than the notion of *intentional* action (Schlosser 2019; Stoutland 2011: 9). Consequently, what distinguishes intentional from non-intentional actions is a central question for both approaches. As Ford points out, the standard approach treats this question (roughly) as a generalised version of Wittgenstein's remark 'what is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?' (1953: §621), the assumption being that the relevant difference is to be found through solving the equation *action* + *x* = *intentional action* (Ford 2011: 77).[13] On this view, what distinguishes intentional from non-intentional actions is an additional, specifiable feature which only intentional actions have. Anscombe, in contrast, rejects the idea that actions are intentional by dint of some additional feature. Such an 'extra feature' (2000: 88), e.g. the agent's entertaining certain beliefs and desires, may of course play a role in our explanation of someone's action. But on Anscombe's view, *it is not in virtue of the subject's having them when she is acting* that the action is intentional: '[w]e do not add anything attaching to the action at the time it is done by describing it as intentional.' (2000: 28).

This is puzzling. For surely there must be *something* which distinguishes intentional from non-intentional actions and to which we appeal when we call an action 'intentional'? At the risk of venturing into a philosophic minefield, let us consider Frege's problem of the concept *horse* to prefigure Anscombe's answer to this question. Frege's semantics for (first-level monadic) predicates, e.g. 'ξ plays the clarinet' and proper names, e.g. 'Ludwig', assigned each of them a semantic value which he called its *Bedeutung*.[14] Frege (1892a) called the *Bedeutung* of a predicate a 'concept' (*Begriff*) and that of a name an 'object' (*Gegenstand*). The concept-object distinction was important for Frege because he wanted the two to play complementary but different semantic roles in determining the truth-condition of sentences containing predicates and proper names. Very roughly, in Frege's view, the proposition – what Frege (1918) called a 'thought' (*Gedanke*) – expressed by 'Ludwig plays the clarinet' is true whenever the object which is the *Bedeutung* of the name 'Ludwig' falls under the concept which is the *Bedeutung* of the predicate 'ξ plays the clarinet' (Frege 1892a, cf. 1891). If Ludwig Wittgenstein is the object in question, the sentence is true, whereas it is false in the case of Ludwig van Beethoven.

[13] As Ford (2011: 77) notes, Wittgenstein himself arguably regarded his question as misguided.
[14] Frege, of course, famously introduced the notion of *Sinn* in addition to *Bedeutung* (Frege 1892b, 1891-1895). We can ignore this notion here.

Despite the importance of the concept-object distinction, Frege found himself unable to express it explicitly. For to say anything substantial about a concept, it seems, one would use a sentence like 'The concept *horse* is F', where 'F' is some property characterising the concept *horse*. But, as Frege recognised, by his own standards 'the concept *horse*' is a proper name and so designates an object. And since, Frege (1892a) thought, concept and object belong to mutually exclusive categories, this led him to say, paradoxically, that 'the concept *horse* is not a concept' (ibid.: 185. Original emphasis.).

Much has been said about Frege's concept *horse* problem (e.g. Wright 1998, Trueman 2015). Here, it is enough to observe that it seems to be a *problem* only insofar as one thinks that, to grasp the concept-object distinction, one must be able to *say* what differentiates them; that the distinction is only intelligible provided one can explicitly describe a feature which applies to, say, concepts but not to objects (or vice versa). But one might think that, at least in a pre-theoretical context, such a specification is unnecessary to grasp the distinction. To grasp the concept-object distinction, one might say, is simply to know how concept- and object-expressions, i.e. predicates and proper names, combine into a meaningful sentence expressing a thought and to know when such a thought is true given the meaning of the relevant concept- and object-expressions. And to know this is to grasp certain *forms of description* (and their interdependence) embodied in certain kinds of expressions of our language which, when combined, allow us to say something intelligible, true or false, about the world. The possession of this understanding does not turn on whether one can specify a criterion which distinguishes the *Bedeutungen* of predicates and proper names. Rather, whether someone has this understanding will be evident from how she uses these expressions. It does not depend on her theoretical knowledge of semantics.

My aim here is not to interpret Frege, but to sketch a model for two assumptions, anticipated earlier, underlying Anscombe's view. First, as with concepts and objects, Anscombe thinks that understanding the difference between intentional and non-intentional actions does not depend on grasping some 'extra feature' (2000: 88) characterising one but not the other. Rather, in her words, 'the term "intentional" has reference to a *form* of description of events' (2000: 84. Original emphasis). Second, for Anscombe, the notions of intentional and non-intentional action cannot be understood in isolation, such that we could *first* have an understanding of actions generally from which we could *then* isolate the subcategory of intentional action (Ford 2011). Just as one cannot understand the semantic role of a predicate independently of grasping how it combines with a proper name into a proposition, so one cannot grasp what a non-intentional action is independently of grasping what an intentional action is.[15]

In Anscombe's view, an illuminating explanation of (intentional) action must bring out a certain form of description, which cannot be captured by an account of the shape *action + x = intentional action*. For such an account assumes that we can

[15] This analogy misleads insofar as, for Frege, concepts and objects play symmetrical or complementary (albeit different) roles in how they 'hold together' (1892a: 193) a thought, whereas, for Anscombe, our understanding of non-intentional action *derives* from our understanding of intentional action (Stoutland 2011: 9). As far as I know, Frege assigned no such priority to either concepts or objects. Indeed, it is doubtful whether, at the bottom level of analysis, there is any basis for an asymmetrical relationship between the two categories at all (cf. Ramsey 1925).

have an understanding of actions generally prior to grasping the form which characterises intentional actions. However, the form of description which Anscombe has in mind *already* permeates our characterisation of someone's action *as* intentional, i.e. it presupposes its own applicability. As such, it demands a different explanation from an 'extra-feature' account (Anscombe 2000: 83) (this will become clearer below). As with a Fregean concept, it cannot, so to speak, be defined explicitly (e.g. in terms of necessary and sufficient conditions). Rather, it must be brought out through exposing, e.g. 'by means of hints', (Frege 1982a: 182), the kinds of descriptions which we apply to intentional actions.

The formal character of Anscombe's approach makes it less straightforward than the standard approach. However, it also makes it more attractive, I think, for articulating a *general* conception of agency, untainted by preconceptions about particular, e.g. rational vs. non-rational, human vs. non-human, kinds of agency. The difference between such kinds of agency is typically drawn by saying that certain creatures *have* something, e.g. language, thought, beliefs, desires, or self-consciousness, which others do not (see e.g. Davidson 1982). But a *form of description*, by itself, is silent on the psychological and physiological make-up of the kinds of beings to which it can be sensibly ascribed. Given that 'extra features', e.g. beliefs and desires, play, if any, a secondary role in Anscombe's account, there is no principled barrier, as far as the relevant form of description is concerned, to ascribing agency to individuals which lack beliefs, desires or thoughts (Anscombe 2000: 86f.; cf. 68f.).

Let us call the relevant form of description discussed by Anscombe *the intentional order*. Anscombe explicates this order through examining the kinds of descriptions we can apply to intentional actions. To that end, she considers the kinds of answers one can sensibly give to the question 'Why?' asked of someone's actions (in different circumstances). Since the types of answers which a question admits reveal the sense in which this question applies, and since, for Anscombe, intentional actions are those 'to which a certain sense of the question "Why?" is given application' (2000: 9) her method illuminates the intentional order. However, as Rödl (2007: 44) remarks, it doesn't provide a unified account of the intentional order from which the relevant answers could be derived.

I will exploit this fact here and concentrate on certain *aspects* of Anscombe's approach, notably the teleological-explanatory form of the intentional order, the idea of animalising descriptions corresponding to what I shall call *thick teleological explanations*, the notion of a process, the formal character of her approach, and normativity. Expanding on these aspects and discussions of them, specifically by M. Thompson (2008) and Boyle and Lavin (2010), I develop a conception of what I call a *productive process*. Finally, to clarify this idea, we relate it to M. Thompson's notion of a *life-form* and Foot's (2001) connected idea of *natural goodness*. The resulting conception will allow us to specify two general constraints on any conception of agency, which will scaffold our discussion in Ch. 3. The chief goal of the present chapter is the motivation of these constraints, not the interpretation of Anscombe. Nonetheless, our discussion should, I hope, illuminate certain aspects of her account.

2.2 Teleology and Explanation

According to Anscombe, 'events are typically described in [the intentional order] when "in order to" or "because" (in one sense) is attached to their description.' (2000: 85). To adopt and adapt her example (2000: 35f.), imagine seeing Beth going upstairs and asking 'Why are you going upstairs?' and she says 'To get my camera'. Here, that *Beth is getting her camera* explains why *Beth is going upstairs*. That is, it is an adequate explanation of Beth's action, i.e. going upstairs, to say that she is doing it *because* she is getting her camera. Initially, this explanation seems strange. For Beth has not gotten her camera *yet*, i.e. her having gotten it is an unrealised future state. But how could something future explain what's going on currently? What is the relevant sense of 'because' here?

2.2.1 The Intentional Order

Let us address the second question first. The relevant sense of 'because' is that of a characteristically *teleological*, i.e. goal-directed or purpose-invoking, kind of explanation. It is the sort of explanation we use when saying that someone is doing A *in order to* or *in view of* doing B (cf. M. Thompson 2008: 89). As Anscombe indicates, such a teleological explanation doesn't always work. If I start waving Beth's camera into her face, her answer 'To get my camera' to my question loses its explanatory force. If, eyes set on the camera, she insisted that she is going upstairs to fetch it, she would 'begin to be unintelligible' (Anscombe 2000: 36). The reason lies in the structure of teleological explanation. For such explanations only work when, as Anscombe puts it, we can see 'how the future state of affairs Q [e.g. Beth's having gotten her camera] is supposed to be a possible later stage in proceedings of which the action P [e.g. Beth's going upstairs] is an earlier stage' (ibid.). More generally, a teleological explanation of the form 'S is doing A *because* S is doing B' or 'S is doing A *in order to* do B' only makes sense insofar as A can be regarded as an earlier stage in a process which leads to S's having done B. When Beth knows that I have her camera, her answer ceases to explain what she is doing because her going upstairs can no longer be regarded as an earlier stage in a process which results in her getting the camera.

Now, suppose Beth's camera is actually upstairs and we ask 'Why are you getting your camera?'. She replies 'To photograph the squirrel in the garden'. Again, Beth's answer that *she is photographing the squirrel in the garden* seems to explain why she is getting her camera. Notice that, because it does so, it also explains, in those circumstances, why she is going upstairs. That is, it seems illuminating to say that Beth is going upstairs *because* she is photographing the squirrel in the garden. Indeed, it seems that we can *describe* Beth's current activity in three ways – going upstairs, getting her camera, and photographing the squirrel – where each description 'swallows up' the former, as Anscombe (2000: 46) puts it. That is, in those circumstances, Beth's going upstairs *is* her getting the camera, and her getting the camera *is* her photographing the squirrel. By the same token, her going upstairs *is* her photographing the squirrel.

But can we really say that Beth's going upstairs *is* her photographing the squirrel, or that she is photographing the squirrel *in* going upstairs? Is going upstairs

really a *stage* of photographing the squirrel? Such doubts point to what Anscombe calls 'the break' (2000: 38) in a series of answers to the question 'Why?' asked of someone's actions. That is, there is a limit to the activities one can cite as providing an *unmediated* teleological explanation ('S is doing A *because* S is doing B') of what someone is currently doing. Some activities, e.g. photographing the squirrel (B), seem to be too distant from someone's current action, e.g. getting the camera (A), to provide such an explanation. In those cases, i.e. those beyond 'the break', it is no longer possible to re-describe ('swallow up') what's going on *here and now* (A) in terms of that distant, larger activity (B) (ibid.: 37-40). The distance must be spanned or mediated by making the purposive connection between these activities *explicit*, e.g. by saying 'Beth is going upstairs *in order to* photograph the squirrel'. Where exactly such a break is reached (i.e. at what point an unmediated explanation, and corresponding re-description, becomes unavailable) is not always clear-cut, and depends on the circumstances (ibid.: 39f.; cf. M. Thompson 2008: 132). I will assume, in our case, that we *can* make sense of re-describing Beth's going upstairs, here and now, as *Beth is photographing the squirrel*. The reader need not agree. The important point is that, in making this assumption, we are supposing that Beth's photographing the squirrel lies *before* the break with regards to her going upstairs.

In sum, processes described in the intentional order 'constitute a series of means' (Anscombe 2000: 47) directed towards a certain end. Because of this, each step in the series can be teleologically explained by its succeeding step(s) and, to that extent, can be truly re-described ('swallowed up') in terms of them (until 'the break'). Thus let us provisionally say: *a process A falls under the intentional order only if A admits of a teleological explanation in terms of some other process B, where A admits of a teleological explanation in terms of B only if A is an earlier stage in a process of which B is a possible later stage.* This characterisation indicates the teleological-explanatory *form* of processes described in the intentional order: roughly, a process described in this form is subject to a teleological explanation by being an earlier stage in a process whose later stages it can bring about. I will call the description of a happening in this form an *intentional description*.

2.2.2 Thick Teleological Explanations

According to Anscombe, 'there are many descriptions of happenings which are directly dependent on our possessing the *form* of description of intentional actions [i.e. the teleological-explanatory *form* of intentional descriptions]' (2000: 84. Original emphasis). For example, when we describe a happening as a *kicking*, *picking up*, or *holding* of something, we depict it as subject to the question 'Why?'. For *kicking*, *picking up*, and *holding* are descriptions of something done as a means to something else which it brings about, and thus as a happening which can be teleologically explained by some end for the sake of which it occurs. These descriptions would have *no sense*, Anscombe thinks, if we didn't understand the form of teleological explanation characteristic of intentional descriptions. I will call descriptions whose sense is so dependent on the teleological-explanatory form of intentional descriptions likewise 'intentional descriptions'.

Now, many things can be described as *picking up* or *holding* something. However, Anscombe sees a difference within the class of such intentional descriptions:

> A dog's curled tail might have something stuck in it, but that of itself would not makes us speak of the dog as holding the object with its tail; but if he has taken between his teeth and kept there some moderate-sized object, he is holding it. To speak of the wind as picking things up and putting them down […] is to animalize it in our language, and so also if we speak of a cleft in rocks as holding something; though not if we speak of something as held there by the cleft. Trees, we may say, drop their leaves or their fruit (as cows drop calves); this is because they are living organisms (we should never speak of a tap as dropping its drips of water), but means no more to us than that the leaves or fruit drop off them. (2000: 86).

Intentional descriptions of things like wind, rock clefts, and water taps *animalise* what they describe whereas intentional descriptions of trees and cows do not. But what's the difference between saying that trees drop their leaves or that a dog is holding a bone between his teeth versus saying that the tap drops its water or that the cleft is holding a rock? Why are the latter animalising descriptions but not the former? One way to understand Anscombe is the following: intentional descriptions of trees (and dogs) are non-animalising *because they are living organisms*, whereas those of rocks and wind *are* animalising because they are inanimate. But why are intentional descriptions of living organisms not animalising? What do they have that wind, rocks, and taps don't?[16]

Here is a thought. Granted, both inanimate and animate subjects can be intentionally described and thus what they are doing can be depicted as subject to teleological explanation: as something done for the sake of something else. However, only with living organisms can the teleological explanation take on what I will call a *thick* sense. A thick teleological explanation not only explains what the subject is doing by relating it to something else which it brings about. It also explains what the subject is doing as a happening of which the subject itself is the *source* or *cause*; as something *originating from the subject*. If we describe a happening broadly as a 'movement', we can also say, with Aristotle, that thick teleological explanations characterise their subjects as *self-movers* (cf. Boyle and Lavin 2010: 176).

[16] On an alternative interpretation, Anscombe sees the relevant division as grounded in the distinction between *animals* and non-animals, including some organisms, e.g. trees (where animals are distinguished by having *sensation* and *appetite* (2000: 68; 86)). This is suggested by her comment that, although we 'may say' that trees drop their leaves it 'means no more to us' than that the leaves drop off them. But this interpretation doesn't explain why Anscombe says that we *may* say that trees drop their leaves *because they are living organisms*, but we should *never* say that a tap drops it water. For this comment suggests that Anscombe thinks that the difference between living and non-living things (trees and taps) *matters* for whether an intentional description is *adequate* or not. If we understand '(in)adequate' here as '(non)animalising', we can explain why Anscombe thinks the tap-description is inadequate (i.e. because the tap is not *alive* (animate)) whilst *still* acknowledging that there is a difference, however drawn, between animals and living organisms generally. On this (i.e. my) interpretation, this latter difference is just not what grounds the distinction between animalising and non-animalising intentional descriptions. On the alternative interpretation, in contrast, it's not clear why the intentional description of a tree should be any *less* inadequate than the tap-description (since both aren't animals).

Let us call a process subject to a *thick* teleological explanation a *productive* process. A productive process not only occurs for the sake of bringing about something else, it also involves its subject as the *source* or *cause* of what it brings about. A thick teleological explanation, in Boyle and Lavin's words, 'represents its subject as engaged in a process that can explain its *own* realization' (2010: 177. My emphasis). To distinguish between animalising and non-animalising intentional descriptions, we must inquire into the nature of a productive process. What makes a process productive? Which processes are subject to *thick* teleological explanations?

2.3 Productive Processes

First, we should delineate more precisely what we mean by a *process*. Consider the following descriptive schema.

(P) S is doing A

Instances of (P) e.g. 'Beth is getting her camera', 'the bird is landing', 'the burn is filling the loch' describe a *process-in-progress* and express a *progressive judgement.* In Boyle and Lavin's words, instances of (P) typically 'describe the here and now by relating it to a possible future situation' (2010: 176): S's having A-ed. To describe Beth as getting her camera when all she is doing *right now* is going upstairs is to relate what she is currently doing to a future state of affairs: her having gotten the camera.

Following M. Thompson (2008: 122-128), we can sharpen the relevant notion of process (in-progress) through the distinction between *perfective* and *imperfective aspect*. Consider the following statements.

(1) The burn was filling the loch. [imperfective aspect]
(2) The burn is filling the loch. [imperfective aspect]
(3) The burn has filled the loch. [perfective aspect]

Here, both (1) and (2) describe as *incomplete* or *in progress* what (3) describes as completed. Put differently, *the burn's filling the loch*, whether described as taking place in (1) the past or (2) the present, is the incomplete version of a process completed with (3) *the burn's having filled the loch*. As such, (1) and (2) describe a process as *anticipating* or *tending towards* what (3) describes as its end. Observe, however, as M. Thompson (2008: 124) emphasises, that the truth of (1) is compatible with the falsity of (2) and (3). It may be true that the burn *was* filling the loch even if it *is no longer filling* the loch or has *not yet* or indeed may *never* fill the loch. Conversely, the truth of (3) is incompatible with the truth of (2) and implies the truth of (1), i.e. if the loch *has been* filled, the burn cannot *be* filling it, but it must be true that it *was*, at some point, filling it. This logical space of compatibility and implication between trios of statements of the form (1) – (3) underwrites the contrast between perfective and imperfective aspect. This contrast, in turn, captures the idea of a *process* as something inherently future-oriented and 'completable'. Instances of (P) will typically sustain the contrast between

imperfective and perfective aspect, including our descriptions ‘Beth is going upstairs’, ‘Beth is getting the camera’ etc.

2.3.1 A First Criterion

Given this specification of a process, let us ask again: what makes a process productive? First, notice that a process, *S is doing A*, is something internally related to its end state, i.e. the reference to its completion or ‘perfection’, S’s having A-ed, is part of what individuates it as the process that it is. Thus *any* process is subject to a teleological explanation in the following *thin* sense: *it is the process that it is* (under that description) only because it tends towards the realisation of its end. A process is by its very nature a *telic*, *progressive* being whose end is integral to what it is. In this thin sense, to describe anything as a process is to comprehend it through the teleological-explanatory *form* of the intentional order (cf. Rödl 2007: 28-32).

A *productive* process, however, must admit of a *thick* teleological explanation. It must not only (a) be a means to the completion of its end but also (b) characterise its subject as the *source* or *cause* of the process’ completion. How might we distinguish processes which satisfy (a) *and* (b) from those which don’t? Following Boyle and Lavin (2010: 176ff.), we might propose the following criterion: a process, *S is doing A*, is *productive* – what they call a ‘goal-directed activity’ (ibid.) – *iff* it can be employed on the right side of explanations of the form

(PP) S is doing A* because S is doing A

That is, in Boyle and Lavin’s words, a process is productive (in our terminology) if and only ‘if it can figure as the explainer of a less embracing progressive with the same subject’ (2010: 177). Importantly, according to this criterion, a process is productive not in virtue of being explainable by something future, but by itself *explaining* another process. Furthermore, a productive process not only explains any arbitrary selection of processes. It explains only those which are a means to, and occur for the sake of, completing it. Thus, if – following M. Thompson (2008: 112) – we call the collection of processes which are a means to the completion of a larger process the latter’s *parts*, we might also describe a productive process as a process which is ‘the “cause” of its own parts’ (ibid.).[17] But in what sense of ‘cause’ exactly? Before discussing this, a few remarks are in order.

As Boyle and Lavin say, ‘S is doing A*’ must be a *less embracing* progressive. That’s because a process, S is doing A, must be somehow *beyond* the process, S is doing A*, which it explains. We cannot say (in one context) that Beth is photographing the squirrel because she is getting her camera or that Alfi is painting the fence because he is constructing it. The things for the sake of which a subject does something must be ‘at a distance’, yet to be reached, from what they are currently doing (Anscombe 2000: 79). This doesn’t mean that the more embracing processes on the right side of the (PP)-schema have to ‘wait in line’ such that *S is doing A** must be *completed* before they can begin (Beth’s going upstairs *is* her getting the camera, even though the latter description is more embracing and

[17] M. Thompson (2008: 112) is there talking about intentional action, but the point applies more widely (Boyle and Lavin 2010: 177).

Beth has not *gone* upstairs yet). It means that the distance between them, however miniscule, must leave enough logical space to span the contrast between imperfective and perfective aspect, i.e. it must be possible to say that, *in* doing A*, S is *on her way* to doing A, but has *not yet done* A. Otherwise, there is no room for the idea that S's doing A* occurs for the sake of completing A.

However, we must admit one limiting case, namely where a process literally explains *itself as a whole*, i.e. where the relata of the (PP)-schema are identical. Here, our understanding of processes, as something inherently incomplete (spanning the imperfective-perfective contrast) hits a wall. As Frege might say, 'here we are confronted by an awkwardness of language' (1892a: 185). For we cannot say: 'S is doing A because S is doing A' whilst maintaining the imperfective-perfective contrast. But as soon as this contrast collapses, we tend to think of a process in a different register, as something finished, ready to put on the shelf, as inhabiting the order of *being*, not *becoming*. A completed process in this sense is an abstract object of contemplation, something which many subjects can do and have done at different times. We shouldn't be misled, however, into thinking that, therefore, such a process cannot explain itself in the same way in which it can explain its lesser parts. Indeed, for the idea of a productive process to be intelligible, it must be *conceivable* for a process to wholly complete itself, to envision a point where it has explained all its parts, and the only 'part' left is itself. Although language puts a barrier before expressing this point progressively, in a way which preserves the *kinetic, unfinished* nature of processes, its importance for the notion of a productive process is, I hope, clear enough. If we admit this limiting case and, relatedly, admit the notion of a process' being its own part, then any productive process, according to our criterion, can not only explain its lesser parts, it can also explain *itself as a whole*. Put differently, any process, S is doing A*, which is part of a productive process will admit of an explanation in terms of the process, S is doing A, which represents the completion of this whole (where, as a limiting case, ◇ (S is doing A* = S is doing A)).

2.3.2 Are We Asking the Wrong Question?

Let's test our criterion. *Intuitively*, it yields good results. It rules out 'animalising' intentional descriptions of processes whose subject, intuitively, is not the *source* of the process' completion. For example, in the (PP)-instance 'the burn is flowing downstream because the burn is filling the loch', that *the burn is filling the loch* seems incapable of explaining why *the burn is flowing downstream* even though the burn's flowing downstream is a means of filling the loch. So, according to our criterion, *the burn is filling the loch* is not a productive process because it cannot figure as an *explanans* on the right side of the (PP)-schema. Similarly, *the wind is picking up the leaves* is not a productive process because, intuitively, it cannot explain another process, e.g. the wind is blowing, which is a lesser part of the wind's picking up the leaves. It seems wrong to say that the wind is blowing *because* it is picking up the leaves (cf. Boyle and Lavin 2010: 176f.). Furthermore, our criterion captures intentional descriptions of processes which, intuitively, do admit of a thick teleological explanation. When we explain why *Beth is going upstairs* by saying 'because *she is getting her camera*' we are not only depicting Beth as doing one

thing (going upstairs) in order to do another (getting her camera). We also represent Beth as being herself the active source of what is happening. Similarly, *the tree is dropping its leaves* is a process capable of explaining its lesser parts (e.g. the tree is reducing its auxin-production) in a thick teleological sense. Thus, it represents the tree as engaged in 'a process that can be the cause of its own coming to be' (Boyle and Lavin 2010: 177).

With this in mind, we might provide the following characterisation of a productive process:

> A process, *S is doing A*, is productive *iff*
>
> (i) *S is doing A* admits of a thick teleological explanation in terms of some other process, *S is doing B*. [where, as a limiting case, ◇ (*S is doing A = S is doing B*)]
> (ii) Where *S is doing A* admits of a thick teleological explanation in terms of *S is doing B* only if (a) *S is doing A* is a lesser or identical stage of *S is doing B* and (b) *S is doing B* is capable of *explaining* a lesser or identical process with the same subject, *S is doing C* [where, as a limiting case, ◇ (*S is doing C = S is doing B = S is doing A*)]

Unfortunately, this characterisation presupposes an intuitive grasp of thick teleological explanations rather than explaining it. To see this, suppose Alfi asks 'Why is "the river is flowing downstream because it is filling the loch" not a thick teleological explanation? Why isn't it a "productive process" as you call it?' – We answer: 'Because 'the river is filling the loch' cannot figure on the right side of the (PP)-schema' – Alfi: 'But *why* can it not figure there? After all, (PP) is just a *schema* for a *form* of explanation, just as 'ξ plays the clarinet' is just a schema for a *form* of description. Cannot *any* two processes fit this schema as long as *S is doing A* is a lesser (or identical) stage of *S is doing B*?' – Answer: 'Yes, but only in *some* cases can the relevant process be said to *originate* from the subject, to account for its *own* realisation' – Alfi: 'But in which cases?'

What can we answer, besides the question-begging 'In those cases where the "because" of the (PP)-schema expresses a *thick* teleological explanation'? Where have we gone wrong?

Anscombe would probably reply that we are asking the wrong question. Her reason, already anticipated earlier, comes out in the following passage:

> The description of what we are interested in is a type of description that would not exist if our question 'Why?' did not. It is not that certain things, namely the movements of humans, are for some undiscovered reason subject to the question 'Why?' So too, it is not just that certain appearances of chalk on blackboard are subject to the question 'What does it say?' It is of a word or sentence that we ask 'What does it say?'; and the description of something as a word or a sentence at all could not occur prior to the fact that words or sentences have meaning. So the description of something as a human action could not occur prior to the existence of the question 'Why?', simply as a kind of utterance by which we were *then* obscurely prompted to address the

> question. This is why I did not attempt […] to say *why* certain things should be subject to this question. (Anscombe 2000: 83. Original emphasis).

Anscombe is here talking about intentional action, but insofar as these are productive processes (self-movements) we can rephrase her point as follows: it is not that certain processes, namely 'productive processes', can for some undiscovered reason figure on the right side of the (PP)-schema. The description of something as a productive process could not occur *prior* to our understanding of the explanatory nexus inscribed in the 'because' of the (PP)-schema, simply as a kind of description by which we were *then* obscurely prompted to ask 'Why is this process able to explain its lesser parts?'. Put differently, 'productive process' is the description of something which *as such* explains the phases leading to its completion, and which cannot be apprehended apart from or prior to an understanding of the teleological-explanatory nexus which unites its phases into an end-directed whole. So, to ask '*Why* can this process play this explanatory role?' is like picking walnut ice-cream and asking why it tastes like walnut. The identification of something *as walnut ice-cream* excludes the intelligibility of the question.

Alright, suppose we cannot ask *why* a process can figure on the right side of the (PP)-schema. Still, we want to understand when a thick teleological explanation applies. And to understand that, we must grasp the sense of *cause* which the 'because' of the (PP)-schema expresses. What is the *explanatory nexus* connecting 'S is doing A *because* S is doing B'? In what sense is a productive process *cause* of its own parts? Boyle and Lavin trivially note that it is 'whatever sense of cause is implied by the "because" in such explanations' (2010: 177). According to M. Thompson, 'the intended notion of "cause" is not pre-conceived, but is that captured by the "because" of rationalization, whatever it may be' (2008: 112) where 'the notion of rationalization […] is restricted to such as *can* be given a final-clausal or purposive or "instrumental" or "teleological" formulation' (2008: 89. Original emphasis). If we follow M. Thompson, then, it appears, any description 'S is doing A *because* S is doing B' which can be 'transposed' (ibid.: Fn. 7) into 'S is doing A *in order to* do B' expresses the 'because' of the (PP)-schema. But how can we tell which explanations can be so transposed? *Intuitively* 'Beth is going upstairs because she is getting her camera' can, and also 'the cat is crouching because it is stalking that bird' (cf. Anscombe 2000: 86), whereas 'The burn is flowing downstream because it is filling the loch' cannot. But *why* not?

'That is a silly question', Anscombe might say. 'You are presupposing your understanding of the relevant explanatory nexus in any description of what you call a "productive process". Processes so described cannot be understood independently of the explanatory form inscribed in their progressive, end-directed structure. Indeed, you are only interested in the burn and the loch because you apprehend their relation through this explanatory form (cf. 2000: 83f.). So stop asking *why* certain processes are subject to this kind of explanation, and enjoy your ice-cream.'

But this reply misses the point. For, again, in inquiring into the notion of a *productive* process, we want to understand the notion of *thick* teleological explanation, i.e. a kind of purposive explanation in which the subject itself figures as the cause of the process' completion. The question *why* thick teleological explanations apply to certain processes but not to others is not silly. If it was, the

distinction between animalising and non-animalising intentional descriptions would be silly. But it clearly isn't: it tracks a fundamental distinction between living and non-living subjects, between intentional descriptions of processes which are *self*-movements, e.g. the cat is stalking a bird, and those which are not e.g. the wind is blowing the leaves.

Our discussion suggests that the (PP)-schema itself cannot mark this distinction if the explanatory nexus of its 'because' is understood in the thin sense of 'in order to'. To isolate productive processes, it seems, the (PP)-schema's notion of *explanation* invoked in (ii/b) must presuppose a *thicker* understanding of the sense in which a process can teleologically explain its own parts, one in which the *subject itself* is the *cause* of what's happening. This thick sense of explanation or 'cause' is not captured by M. Thompson's notion of rationalisation and it is presupposed, not explained, by Boyle and Lavin's employment of the (PP)-schema. We thus return to our question: what makes a process productive? Perhaps a more promising approach is to consider the *subjects* of productive processes. How does 'S' in '*S is doing A* because *S is doing B*' have to be for the 'because' to express a thick teleological explanation?

2.4 Productive Subjects

2.4.1 The Possibility of Progressive Judgements

Let us begin by asking: how are progressive judgements, i.e. instances of the (P)-schema 'S is doing A' intelligible at all? As explained, progressive judgements, e.g. 'the burn is filling the loch', 'Beth is getting her camera', may be true even if S has not yet completed A and perhaps never will. This is quite puzzling. For suppose that Beth's camera has been stolen by extra-terrestrial secret agents, so she will never get it. How can it then be true to say, here and now, as she's going upstairs, that Beth is getting her camera?

Following Boyle and Lavin (2010: 178ff.), to answer this question we need to observe that progressive judgements do not relate a subject's activity to what will *really* happen, but to what will *normally* happen or *tends* to happen *if nothing interferes*. In their words, progressive judgements don't make reference 'to *the actual* future [of the subject], but rather to *its own* future' (ibid.: 180. Original emphasis). To say that Beth is getting her camera is not to say that she *will* definitely get it. It is to say that she is on her way to getting it and thus will get it if nothing interferes. But how can we distinguish between interfering and non-interfering futures? What determines an outcome as a subject's *own* future?

Following Aristotle, Boyle and Lavin (ibid.) propose the following answer. To be intelligible, they say, progressive judgements presuppose the truth of two further kinds of judgements, what they call *judgements of form-attribution* (FA) and *form-characterising judgements* (FC):

(FA) S is an *F*
(FC) *F*s do α (in conditions C)

FA-judgements ascribe a certain general *form* or *kind* or *nature* to a subject, understood in broadly Aristotelian terms as a principle of motion and change (ibid.: 179). These forms need not be species-concepts such as *human* or *puffin.* They may also include items such as *photographer*, *German*, *sister, burn* or *glacier*; that is, generally speaking, items capable of being the subject of an FC-judgement. FC-judgements correspond to the linguistic category of *generics*, i.e. unquantified generalising descriptions ascribing properties to a kind without determining how many of its members exhibit the predicated property, e.g. 'photographers are introverts', 'Germans work efficiently', or 'glaciers melt (in the sun)' (Leslie and Lerner 2022). (FC) focusses on the bare plural form 'Fs', but generics can also be expressed with a singular subject, e.g. *the tick* carries Lyme disease, *SARS-Cov-2* causes COVID-19. The semantics of generics is tricky. Here, it's enough to emphasise that they express *general* judgements concerning properties of things of a certain nature, kind or form.

By considering progressive judgements about an individual's activity against the background of FA- and FC-judgements, we can make the idea of (non)interference intelligible. Suppose the FA-judgement *Beth is a photographer* and the FC-judgement *photographers take pictures of salient objects (in certain circumstances)* are true. Then, given the circumstances (a squirrel in the garden, a camera upstairs), that *Beth will get her camera* is what we expect will happen. Conversely, if she does not end up getting it, we would be surprised. Did she trip? Did the squirrel disappear? Indeed, even if any of these situations occurred, it would still have been true to say, back then, as Beth was going upstairs, that *she is getting the camera*. Given the truth of our two judgements, her getting it is what *tends* to happen, and thus her *not* getting it, because it contradicts this tendency, counts as an interference.

Of course, there may be many FA- and FC-judgements true of an individual which underlie our progressive judgements about what's going on e.g. *Beth is human*, *humans walk upstairs without tripping (in certain conditions)*, all of which inform our expectations of what *tends* to happen. Furthermore, whether and why a particular FC-judgement is true and, as Boyle and Lavin (2010: 180) note, whether S's doing A is an instance of *α-ing* may be difficult to determine. Is Beth really going after the squirrel? She said so, but what if she was lying? Finally, it seems implausible that we consciously make such judgements whenever we describe a process-in-progress. So how illuminating is this account?

Boyle and Lavin's point is not that certain FA- and FC-judgements are clearly true of any subject currently engaged in a process, nor do they intend to describe the cognitive mechanisms underlying such judgements. Their proposal is interesting because it characterises a general structure of *forms* of judgement and their relations which illuminate how progressive judgements are intelligible at all. It formally captures the Aristotelian thought that to judge what a subject is, here and now, doing, we must place its activity within the context of the subject's general nature or form. For insofar as this form specifies general principles of what subjects of this form do (in certain circumstances), i.e. general tendencies of movement and change, it makes the idea of a subject's *own* future, of what will happen if nothing interferes, intelligible. Thereby it explains how it can be true, here and now, to say that *S is doing A* even though S hasn't *yet done* A and may *never* (completely) do it.

Without such general conceptions of the form of subjects underlying our progressive judgements, it would be hard to see how progressive judgements are even possible. After all, we don't *know* the future. Hence, if we had no idea about the general kinetic tendencies of a subject, then all we could say about a subject's current activity is 'S is doing what will happen'. But then absolutely *anything* (within the bounds of logic) could count as 'what will happen' and it would seem arbitrary to single out one of the million different possible futures as indicating the correct description of what S is, here and now, doing. Thus, Boyle and Lavin's proposal not only explains how progressive judgements can be true, it also contains a transcendental insight: it articulates the conditions of possibility for the intelligibility of progressive judgements.

2.4.2 Isolating Productive Processes

The above point holds not only for judgements concerning *productive* processes, e.g. *Beth is getting her camera,* but for processes per se. For instance, that *the rock is falling to the ground* likewise presupposes the truth of (FA) *this is a rock* and (FC) *rocks fall to the ground (in certain circumstances)*. Thus, the proposal, despite illuminating the possibility of progressive judgements, cuts across cases in which a subject's 'own future' itself explains its current activity, in a *thick* teleological sense, and those in which it does not. To mark this distinction, we must explain, in Boyle and Lavin's words, 'not just how a certain future can be a subject's own, but how it can be so in a way that explains the subject's present activity.' (2010: 181).

Following Boyle and Lavin (2010: 182), here is a promising thought: suppose that what a process, *S is doing A*, is tending towards, i.e. S's having A-ed, is itself a *precondition* for S's capacity to do A at all. Then, it would be intelligible how the subject's own future, i.e. S's having A-ed, can explain S's doing A now in a way which depicts S as the *source* of doing A. For if the completion of A is itself a precondition for S's doing A, which in turn is a process tending towards A's completion, then we get a grip on the idea of a process which accounts for its *own* realisation. But how can a process be its own precondition?

Boyle and Lavin's (ibid.) answer is: by being part and parcel of a *system* of processes on which the process itself depends. On this proposal, a process, *S is doing A*, is its own precondition just in case its effect, S's having A-ed, itself contributes to sustaining a system of processes on which S's doing A depends. A process A belonging to such a system will always lead back to itself in that it will precondition some other process, which preconditions yet another process, and so on, until we eventually reach a process which is a precondition for A. Such a process A *enables itself* because what it brings about can be regarded as 'an effect whose coming to be contributes to the very conditions that make its own coming to be no accident' (ibid.).[18]

This idea of a self-enabling system of processes is quite abstract. However, given our transcendental insight above, we can already stress that the relevant system must consist of processes which specify *general tendencies* characterising the subject's form, explaining what subjects of this form generally do. Otherwise we could not understand how a subject engaged in a process occurring *here and*

[18] Notice the similarity to precariousness (Ch. 1). We examine this connection in Ch. 3.

now is related to a system of processes (which may not be occurring now) on which it depends and to which it contributes. We can envision the generality of such a system by taking it to be constituted by processes described in FC-judgements. Relatedly, we can think of it as constituted by the kind of process we discovered as a limiting case of the (PP)-schema. That is, as a process which can be instantiated at different times in different places by different subjects.[19] I will call such a process a *generic process*.

Given the idea of a self-enabling network of generic processes, let us sharpen our notion of a productive process:

A process, *S is doing A*, is productive *iff*

(i) *S is doing A* admits of a thick teleological explanation in terms of some other process, *S is doing B*. [where, as a limiting case, ◇ (*S is doing A = S is doing B*)]
(ii) Where *S is doing A* admits of a thick teleological explanation in terms of *S is doing B iff*
(a) *S is doing A* is a lesser or identical stage of *S is doing B*
(b) S's form is a self-enabling network of generic processes and
(c) *S is doing B* is an instance of a generic process in this self-enabling network

A process which satisfies (i) and (ii) meets our earlier criterion, i.e. it will be capable of occupying the right side of the (PP)-schema. However, unlike that criterion, the present definition does not presuppose an grasp of thick teleological explanations. Rather, it says that a process admits of such an explanation because it belongs to a certain kind of self-enabling system. Thus understood, our characterisation equips the 'because' of the (PP)-schema with a *thicker* sense of teleological explanation than 'in order to'.

Given this characterisation, we can say that a subject's own future, i.e. it's having A-ed, anticipated in the progressive judgement *S is doing A* can thickly-teleologically explain the subject's doing A *now* just in case *S is doing A* is productive as defined above, i.e. just in case it is itself (part of) a process, *S is doing B*, which instantiates a process belonging to the self-enabling network of generic processes which constitutes S's generic form.

Our characterisation, though abstract, illuminates the distinction between productive and non-productive processes. It excludes explanations such as *the burn is flowing downstream because it fills the loch* not because they *feel* wrong, but because *the burn's filling the loch* is not a process within a self-enabling network of processes which constitutes the general form *burn* and which enables the burn to flow downstream. In other words, *the burn is filling the loch* is not itself a process whose realisation is a precondition for the burn's flowing downstream. In contrast, 'the cat is crouching because it is stalking that bird' is a description of a *productive* process since stalking the bird, among other things, *enables* the cat to crouch. For stalking birds is, generally, a way of catching them is a way of eating them is a way of digesting them is a way of getting energy is a way of enabling the crouching

[19] Rödl (2007: 37) calls such a process an *infinite end*.

needed to catch them. From that perspective, the cat is not only 'stalking a bird *in* crouching' (Anscombe 2000: 86), it's also crouching *in* stalking the bird: its stalking contributes to the conditions which make its crouching no accident. Granted, the enabling relation between stalking and crouching is highly mediated. However, unlike in the case of the burn, we can at least tell a story of how the stalking preconditions the cat's ability to crouch (Boyle and Lavin 2010: 182f.). Our proposal delivers these initially promising results because it makes reference to a certain *generic self-enabling form belonging to the subject* of intentional descriptions. Without reference to this form, we would fall back on our thin understanding of teleological explanation, with only intuitions to decide which kinds of processes admit of *thick* teleological explanations.

In sum, our sharpened characterisation provides a first conception of a productive process as a purposive movement which *originates* from its subject, accounts for its *own* realisation or moves *itself* through being its own precondition. However, the idea of self-enabling form in (ii/b) which this conception involves requires elaboration: we must explain how this form is self-*enabling*, how its constituent processes *precondition* each other. Furthermore, our proposal also counts processes as productive which, intuitively, are not *actions*. *The cat's heart is pumping blood* is a process which enables the organism to function, which, in return, enables the heart to pump blood by contracting. So, 'the cat's heart is contracting because it is pumping blood' is a thick teleological explanation. Intuitively, however, the heart's activity is not an action in the sense in which the cat's crouching is an action. As this suggests, actions are productive processes, but not all productive processes are actions. What is special about actions? We return to this later. Before, however, we must consider one important dimension of the intentional order, and of productive processes in particular: normativity.

2.5 Normativity

As mentioned earlier, certain answers to Anscombe's question 'Why?' are misplaced. Beth's saying 'I'm getting my camera' to my question 'Why are you going upstairs?' when she sees it in my hand is puzzling. However, although her answer is puzzling, the question itself is not misplaced. Indeed, it is because the question is *not* out of place that her answer *is* out of place. To see this, suppose Beth had instead replied 'I wasn't aware I was going upstairs' or that, upon inspection, we find her in a sleepwalking soliloquy. Then the question itself would have been misplaced. Such circumstances constitute what Anscombe calls a 'rejection of the question' (2000: 25). They exclude events from falling under the intentional order, and consequently, prevent the question 'Why?', in its teleological-explanatory sense, from having application. In our initial scenario, Beth does not reject our question because, although her reply 'I'm getting my camera' is strange, what Beth is doing, i.e. going upstairs, seems to be a process carried out for the sake of something else, e.g. getting her camera, through which it may be explained. To this extent, Beth's activity falls under the intentional order. This is why the 'Why?'-question applies and why Beth's answer is unintelligible. Since her camera is in my hand, her reply fails to explain her going upstairs *understood as part of a productive process,* i.e. getting her camera.

Beth's answer is unintelligible in what we can call a *subject-independent* sense. The reason for the inadequacy of her answer has nothing to do specifically with Beth herself. Rather, since her camera is in my hand, we simply cannot see Beth's going upstairs as an earlier stage in proceedings of which her getting the camera is a later stage. Her answer clashes with our understanding of the unfolding, causal nexus of processes quite generally, and this understanding is independent from any particularities of Beth.

There is also a different sort of unintelligibility. Suppose there is a *grey* squirrel in the garden, (unfortunately) an increasingly common sight nowadays. I give Beth her camera and ask 'Why are you getting it?'. She says 'To photograph the squirrel in the garden'. Here, we might find Beth's answer unintelligible in the following way: we do not see *the point* of photographing the squirrel. We don't understand why photographing the squirrel is a goal Beth pursues, given that grey squirrels are boring. Again, this kind of unintelligibility only arises if what Beth is doing, e.g. getting her camera, is understood as a goal-directed activity, as falling under the intentional order. It is because her activity occurs for the sake of some yet unrealised purpose, e.g. photographing the squirrel, that our question '*Why* are you getting the camera?' applies. But the inadequacy of Beth's answer is different from above. It is not that we cannot see her getting the camera as an earlier stage in the larger process of photographing the squirrel. Rather, her answer is unsatisfying because we cannot see why photographing the squirrel is an end *worthy of pursuit*.

As this illustrates, the question 'Why?' not only seeks an answer which places the subject's current activity within a causal nexus terminating in an end she is pursuing. It also asks for what Anscombe (2000: 70ff.) calls a 'desirability characterisation', i.e. a description of what someone is doing which shows it to be pursuit-worthy or good in some respect (cf. Davidson 2001: 3). The question 'Why?' not only demands: explain what the subject is doing by relating it to something else which it can bring about. It also demands: explain in what sense the subject's doing has a *point*, is *pursuit-worthy* or 'to be done' (Rödl 2007: 18). Anscombe's 'Why?'-question is thus a way of asking 'What for? What's the good in it?'. This question appeals to the subject's concerns, or, more generally, it tries to reveal the sense in which what the subject is doing *matters* to it, why it makes sense *for the subject* to pursue some end. A desirability characterisation provides an answer to this element of the 'Why?'-question. It stops any (further) question 'What for?' from arising (Anscombe 2000: 70ff.). Consequently, an answer which lacks a desirability characterisation, i.e. which fails to explain how the subject's activity is pursuit-worthy, gives rise the unintelligibility illustrated by the grey-squirrel-case. We can call this unintelligibility *subject-dependent* because it involves the subject's concerns.

2.5.1 Standards

What counts as a desirability characterisation? Generally, we can say: it must involve the subject's concerns. But how exactly? Consider the following (adapted) example from Anscombe (2000: 62ff.). I catch Beth on her way to the Hereford market and ask 'Why are you going there?'. She says 'To see the Jersey cows'. As it stands, this answer is not a desirability characterisation, i.e. it leaves room for

asking further 'What do you see them for?'. Beth replies 'To see if they are strong enough' – Me again: 'Strong enough for what?' – Beth: 'For my farm'. Here, finally, we see what Beth sees in going to the Hereford market. For her final answer implies that she is a *farmer* and that she needs strong cows for her farm, and it is pointless to ask a farmer 'What do you need a strong cow for?' (ibid.: 66).

Notice what has happened here. Beth's final answer provides a desirability characterisation by placing her activity against the background of *being a farmer*, from the perspective of which going to the Hereford marked (to examine the cows) is *as such* desirable. This suggests the following clue for an adequate desirability characterisation: it identifies the subject's activity as falling under a more general kind, principle or form – let us call this a *standard* – from whose viewpoint its doings are as such pursuit-worthy. Anscombe herself (following Aristotle) seems to have placed something along those lines as a *constraint* on adequate answers to the 'What for?' question given that her discussions of adequate desirability characterisations turn on whether the subject falls under some general kind, e.g. *human* (2000: 64, 72), *farmer* (2000: 70), *Nazi* (2000: 72), and whether, as such, what the subject is doing is good, desirable, to be done etc. We can motivate this constraint by considering the end-point anticipated by a series of 'What for?'-questions. For such a series tends to continue as long as the answers (a) fail to invoke a *general* standard from the perspective of which the subject's activity is pursuit-worthy or (b) fail to show how the subject *falls under* a general standard from the perspective of which the subject's activity is pursuit-worthy. Whereas (a) implies (b), the reverse is not the case. For instance, if Beth's final answer had been 'for *Alfi's* farm', we would have asked further 'What for?' viz. 'Why do you care about Alfi's farm?'; however, not because we cannot see why strong cows are pursuit-worthy for farmers, but because we cannot see how the standard provided by the general kind *farmer* applies to Beth. The missing connection is provided by e.g. 'Alfi is my father' (and *daughters* support their farmer-fathers) or, as in Beth's answer, 'for *my* farm' (and *farmers* care about strong cows). Thus, apart from invoking a general standard, a desirability characterisation must show how this standard *governs* the subject, i.e. why it is subject *to it*. So much for (b). But why, as (a) implies, must the relevant standard be *general*, and in what sense?

A full answer to this question lies beyond our present topic. Here, the following observation should suffice. For a standard to have any normative force, it must, in principle, be possible for an individual to *succeed* or *fail* to accord with it. But to account for the possibility of failure, a standard must be *general* in the following, minimal sense: it must, in principle, be capable of governing the activities of different individuals. It must provide a basis for saying, of different individuals in the same circumstances, that the activity of one accords with the standard but the other does not. Otherwise there is no room for *evaluating* an individual's activities relative to the relevant standard. If a standard 'governed' only the activity of one individual, it would be impossible to say whether or not it's activities accord with it. In such a case, as Wittgenstein remarks in a related context, 'whatever is going to seem right […] is right. And that only means that here we can't talk about "right"' (1953: §258). Such a 'standard' would, in other words, not be a *standard* at all: it would have no normative force, enabling evaluations of an individual's activities, e.g. as good or bad, right or wrong, success or failure. Thus,

a standard must be general in the minimal sense of being capable of governing the activities of different individuals.

2.5.2 Two Constraints on Normativity

Our discussion illustrates that intentional descriptions bring with them a specific kind of normativity: an evaluative dimension. Processes falling under the intentional order not only form a series of means and ends, they are also governed by a standard relative to which they can be *evaluated*, e.g. as good or bad, pursuit-worthy or pointless, (un)desirable, (not) to be done. Our discussion suggests two features of this normativity. First, it is *internal* to the intentional order, i.e. it applies to a process, *S is doing A*, because that process falls under the intentional order (cf. Rödl 2007: 20). For the question 'What for?' sensibly applies only to a process carried out for the sake of something else through which it can be explained. Furthermore, it applies whenever a process admits of such a teleological explanation. Second, the standard on which this normativity rests must satisfy two constraints: it must (a) be *general* and (b) *govern* the subject's activity, i.e. its activity must be subject *to it*. An adequate grounding of the relevant standard should meet these constraints.

We must disarm some preconceptions about the *kind* of normativity and corresponding standard. First, it is not (necessarily) ethical or moral, but rather inherently *practical*. That is, the notion of goodness specified by a desirability characterisation must only explain how the process in question *makes sense* from the subject's point of view. This thin, practical-explanatory notion of 'good' ('desirable', 'pursuit-worthy') is 'multiform', as Anscombe's (2000: 75) says. It can take shapes which are unethical but which nonetheless answer our 'What for?' question, as when to 'Why are you going to the Hereford market?' Beth had replied 'To shoot Alfi's cow, he ran over mine'. Second, *insofar as it is internal to the intentional order,* the relevant normativity can govern processes which are not actions. It is, for instance, sensible to ask of the contraction of a heart 'What for?'. If then one answers 'Pumping blood', one can again ask 'What for?', and this series of questions will continue until we reach a general standard, e.g. the overall health of the organism, which governs the heart's activity and from the perspective of which it is, as such, desirable. This suggests that the relevant notion of desirability (or goodness) can, pace Anscombe, not 'only be ascribed to creatures endowed with sensation' (2000: 68; cf. 86), i.e. animals, her prime examples being humans, dogs (ibid.: 68) and cats (ibid.: 69; 86). It also applies to 'subjects' of processes, e.g. *the heart*, which have no 'psychical' capacities such as sensation, perception, affect, want, desire, beliefs and intentions; subjects which, accordingly, seem to be incapable of *actions*. This is not to deny the difference between the latter kind of subject, i.e. agents, and the former. But it is to resist the thought that this difference marks a break in the *kind* of normativity which is inscribed in the intentional order and which defines the most general sense in which any 'What for?'-question *evaluates* a given (productive) process. As explained below, we can justify this resistance by noticing that the formal relation holding between, say, the heart's activity and its governing standard and Beth's activity and its governing standard

is *the same* and that this formal relation can ground an intelligible notion of desirability or goodness.

2.6 Life-Forms: A Case Study

Let us, finally, examine M. Thompson's notion of a *life-form*, which is intriguing for two reasons. First, it formally accounts for the idea of a self-enabling system of generic processes and of the normativity governing productive processes. Second, the notion involves certain difficulties which reveal the challenges for a naturalistic account of productive processes. Thus, apart from substantiating the idea of a productive process, our discussion should bring out the challenges ahead.

2.6.1 What Is a Life-Form?

Recently, I explored a nearby forest. Jolted by a blow to the head, I turned to fight the culprit, only to find a cone on the moss. Inspecting a nearby tree, and browsing through my *Mitchell Beazley pocket guide to Trees*, I found a matching description: 'Cone 8cm, ovoid, green in 1st year. Scales not spined. Bark in upper crown orange, flaking, heavily fissured at base' under the heading 'Scots pine' (Rushforth 1980: 60). 'Aha!', I said, 'this is a Scots pine' and then, mumbling to myself 'Scots pines have 8cm long, ovoid cones with unspined scales.'

In M. Thompson's view, my out-mumbled thought has a special *logical form*, expressed in statements like 'The S is/has/does F' or 'S's have/are/do F', e.g. 'The Scots pine has orange bark in the upper crown' or 'Scots pines have 8cm long, ovoid cones with unspined scales', as familiar from nature documentaries and field guides. Such statements express what M. Thompson calls a *natural historical judgement* (2008: 64), henceforth 'NHJ'. Here, 'The Scots pine' does not denote a particular organism, e.g. *this* Scots pine over there. It refers to Scots pines *in general*, considered as a species, or, in M. Thompson's words, a *life-form*. M. Thompson argues that the generality of NHJs is irreducible to other kinds of generality, e.g. generic or Fregean (quantifiable) generality. Nor is it statistically analysable, as saying e.g. '*Most* S's are F' or by appeal to constraining 'ceteris paribus' conditions, as saying e.g. 'For all x, if x is a Scots pine, and conditions c_1, …, c_n are satisfied, then x's upper crown bark is orange', where 'c_1, …, c_n' are understood as normal or standard conditions for a given individual organism, or are subjected to a normative reading (M. Thompson 2008: 68-76).

Although these points are disputable, let's suppose, following M. Thompson (1995, 2004, 2008: Pt. I) and Foot (2001), that NHJs, in virtue of being judgements about *living things*, are of a logically special, formally distinctive kind, corresponding to the idea of a *life-form*. But what exactly is a *life-form*? Consider the following NHJs.

(a) In spring, two years after pollination, the pine cones open.
(b) In spring, two years after pollination, the pine's seeds are released.

As Thompson (2008: 65) explains, such claims bear a peculiar timelessness, despite invoking 'spring' and 'after'. We are not talking about *this* spring or about a moment two years *after* some specific past event, i.e. *this* pine's pollination in 1684. Rather, (a) and (b) articulate phases in the pine's life-cycle, a cycle with a certain temporal structure or rhythm, which is somehow 'cut free' from any specific moment – just as a musical rhythm, it specifies tim*ing*, not moments *in* time.

The NHJs articulating phases of a given life-cycle allow for purposive connections (ibid.: 77), e.g. (a) and (b) may be joined by saying that 'In spring, the pine cones open *in order to* release the pine's seeds.' This affinity for teleological connection applies generally to NHJs specifying different parts, aspects, and phases of a given life-form. Thus, taken together, the NHJs true of a given life-form constitute a kind of teleologically closed unity (ibid.: 79, Haase 2018: 93f.). Just as, to paraphrase Ramsey (1925: 410), 'Socrates' determines a set of propositions 'collected together' by sharing the same form ('ϕ Socrates'), so, roughly, a *life-form* determines a set of NHJs collected together by the 'in order to' connective. A particular life-form, then, may be conceived as the way in which a set of true NHJs connect together into a formally distinctive, teleologically closed unity or life-cycle with a certain atemporal structure (Foot 2001: 33f.). This teleologically closed organization of NHJs is what Thompson (2008: 78) calls a life-form's *natural history* (not to be confused with evolutionary history), which determines the 'nature' of the life-form.

According to M. Thompson, *any* description of a process as *vital*, i.e. as a *life*-process, implicitly makes reference to a 'wider context' (ibid.: 53ff.), which is the life-form. It is only in relation to a particular life-form that I can describe a process as playing a role in some organism's life. When I point at a pollen-laden breeze, saying 'look, *pollination* is in progress', I am, to paraphrase M. Thompson (2008: 78), pointing implicitly *at the life-form of the Scots pine*. Remove the form, and all we have is bits flying purposelessly through the air. Put differently, the description of something as *pollination*, *breathing, eating, sex etc.* or as *gills*, *the marmot's worst enemy* or an *individual organism* presupposes the background of a life-form against which what one describes adds up to pollination, breathing, etc. Indeed, as M. Thompson (2008: 55) notes, what is physically the same process, e.g. mitosis, amounts to *reproduction* in one case (e.g. in amoebas) and to *growth* or *self-maintenance* in the other (e.g. in humans).

M. Thompson's position parallels Anscombe's (2000: 80-87) point about intention-dependent concepts (e.g. *holding*), that our understanding of the intentional order pervades and pre-structures our individuation of happenings through the lens of intentional descriptions, revealing 'the order that there is in this chaos' (2000: 80), i.e. the complex goings-on coinciding with someone's doing something. Indeed, echoing Anscombe (2000: 83), we might paraphrase M. Thompson as proposing that the description of something as a living organism (or a life-process) could not occur prior to our understanding of life-forms, simply as a kind of utterance by which we were *then* obscurely prompted to address the question 'Is it alive?'. This Anscombian theme underwrites M. Thompson's scepticism as to whether life admits of a 'real definition' in terms of a (list of) positive feature(s) present in an otherwise neutral description of happenings (2008: 33-48).

2.6.2 Natural Goodness

Life-forms provide a way of explaining observer-independent, evaluative judgements about individual organisms. As Foot (2001: 30), expanding on M. Thompson (2008: 80), explains, a judgement such as 'This puffin is defective' can be explained through the relation between two formally distinctive kinds of judgements: (i) judgements concerning the actual state, process or behaviour of an *individual* organism and (ii) true *general* judgements (NHJs) about its life-form (Foot 2001: 30). That is, if 'The puffin has wings' and '*This* puffin has no wings' are both true, we may infer that something is *wrong* with this particular puffin (M. Thompson 2008: 80). Conversely, 'The puffin eats fish' and '*This* puffin has caught a sandeel' support the judgement that what's happening is *good* (desirable) for this particular puffin. This formal-evaluative relation between an individual and its life-form is the inspiration for Foot's notion of *natural goodness*, i.e. a goodness 'which is attributable only to living things themselves and to their parts, characteristics, and operations [and is] intrinsic or "autonomous" goodness in that it depends directly on the relation of an individual to the "life form" of its species' (2001: 26f.).

M. Thompson claims that the conditions necessary for a life-form specific phenomenon to occur themselves belong to the form's natural history, even if they 'include a feature of the environment' (2008: 79). For example, pollination is required for the Scots pine's cones to open, but pollination relies on wind. Hence, the presence of wind, somewhere in between the cone's opening and pollination, belongs to the natural history of the Scots pine. Such environmental features are, as M. Thompson's writes, '"presupposed" by the life-form itself' (2008: 71). Again, that's because any description of a process as a *life*-process already contains a reference to a life-form, e.g. to describe a breeze (an environmental process) as *pollination* only makes sense in relation to the life-form *the Scots pine* (or other plants).

Let us draw a rough distinction between what we can call the *physical* environment of an individual organism and the environment as part of its life-form, its *natural* environment. The former may be understood, roughly, as the goings-on outside the physical boundary, e.g. the membrane, skin, bark etc. of an individual organism. Whereas the latter is a kind of subset of these goings-on, carved out in virtue of being necessary for (or playing a role in) maintaining the life-cycle of a given life-form (cf. Foot 2001: 33ff.).

If M. Thompson is right, the notion of a *natural* environment is presupposed by the life-form to which it belongs. This does not mean that we consider NHJs specifying features and inner parts of individuals, and then see what's 'left over', to determine its natural environment. The subset of vital processes outside the physical boundaries of an individual organism, i.e. the natural environmental conditions necessary to maintain its life-cycle, are *as much part of* the life-form as are features of its phenotype. Wind *belongs* to the life-form *the Scots pine* just as much as needles, bark, cones and pollen, since it forms part of the causal structure on which the existence of this kind of life depends. As M. Thompson says 'the teleological connective ['in order to'] simply expresses the concept that is converse to this sort of dependence.' (2008: 79). Thus the system of NHJs characterising a life-form will also contain judgements specifying a creature's natural environment,

e.g. ‘Koala’s live in such-and-such climates’ or ‘Bobcats breed in spring’ (ibid.: 71).

This explains M. Thompson’s comment that the notion of natural defect at play in our judgements of natural goodness is ‘is so unnaturally broad that it would take in, say, losing aspects of the individual creature’s environment’ (2008: 80, Fn. 16). Thus picture, say, Betty, the most perfect koala-specimen imaginable. Now place Betty in the Arctic. There seems to be *something wrong* here, normatively speaking. But what? By hypothesis, it isn’t Betty. Nor is it the Arctic as such. We cannot make sense of this kind of evaluative judgement, it seems, by considering the individual organism and its physical environment in isolation from each other. Its possibility resides in the fact that the koala’s *natural* environment, because it is *part of* or *presupposed by* its life-form, is carried into the Arctic scene through the identification of Betty as bearing the life-form *the koala*. The ‘wrongness’ of the situation, then, arises from the incongruence between the physical environment of this particular koala (the Arctic) and the natural environment presupposed by its life-form (warm climates).

As Thompson notes, such evaluative judgements do not, strictly speaking, express the concept of something’s being wrong with this particular organism, but rather of ‘something’s being wrong *in connection with* the organism’ (2008: 80, Fn. 16. My emphasis). However, this is not to deny that the formal-relational structure underlying our evaluative judgements is *the same* in both the koala and the puffin case. Both cases relate judgments of type (i) and (ii): we reason from ‘The S is F’, e.g. the puffin has wings, koalas live in warm climates and ‘This S is not F’, e.g. this puffin has no wings, this koala lives in the Arctic, to ‘there is something wrong (in connection with) this puffin/koala’.

As this suggests, the formal structure underwriting judgements of natural goodness does not itself discriminate between features of the individual and its physical environment. Relatedly, the distinction between ‘something’s being wrong *with the organism*’ and ‘something’s being wrong *in connection with* the organism’ is not a logical distinction, one intelligible at the level of life-*form*. It seems to presuppose the grasp of a further, concrete and empirical distinction, namely that between *this* individual organism and *its* physical environment. From the perspective of the life-form itself, there is no basis for this distinction. All that matters to the notion of a life-form is the existence of a formally distinctive, teleologically closed unity of NHJs, and we have seen that the NHJs constituting this unity cut across phenotypic aspects of the individual, its activities, and its environment. Loosely put, a life-form is just a bunch of purposively connected processes stitched together by sharing a certain kind of general form. Indeed, on M. Thompson’s picture, it is only *because* these processes share this form that they count as *life*-processes at all. But which of these processes concern the creature’s environment and which do not is irrelevant to the notion of a life-form. At the level of form, the *concrete individual organism*, and with it the distinction between NHJs concerning the environment (i.e. those specifying its natural environment) and NHJs concerning parts, features and activities of the individual, dissolves.

2.7 Life-Forms and Beyond

2.7.1 The Significance of Life-Forms

Life-forms help to substantiate the notion of a self-enabling system of generic processes underwriting our characterisation of a productive process. The NHJs constituting a life-form formally match what we called *generic processes*. First, like FC-judgements, NHJs specify what beings of a certain kind *generally* tend to do or be like (in certain circumstances). Second, they exhibit the same atemporality as generic processes (i.e. the repeatability marking the limiting case of a completed productive process).

Life-forms also seem to be *self-enabling* systems in that their generic constituent processes mutually depend on each other. Indeed, as Foot (2001: 30ff.) stresses, an essential feature of life-forms is that the activities and properties specified by NHJs all relate to the self-maintenance, development and reproduction of the relevant life-cycle. In Kant's words, in living, organised beings 'nothing […] is in vain, purposeless, or to be ascribed to a blind mechanism of nature' (2000: 248). This non-accidental, mutual dependence of life-processes is what, in Foot's view, grounds the teleology internal to life-forms. It's because life-processes form an interdependent, self-sustaining system, in which every process *matters* for realising the whole, that a particular process in this system can be said to occur *in order to* maintain another (Foot 2001: 32ff.; cf. Boyle and Lavin 2010: 182f.).

Life-forms also illustrate the intimate connection between such purposefully connected, self-sustaining networks and normativity. Life-forms provide a standard which satisfies our two constraints on the normativity implicit in intentional descriptions. They are (a) *general* thanks to their generic form, and thus provide a standard which applies equally to different individuals (those which bear the relevant life-form). Furthermore, life-forms (b) *govern* the activities and features of the organisms which instantiate them. This follows directly from M. Thompson's account since he defines an *individual living organism* as 'whatever falls under a species or "bears" a life-form' (2008: 76f.). If an individual organism *is* the bearer of a life-form, then its activities are bound by the standard supplied by its form. It cannot *not* uphold the generic processes constituting its life-form any more than it cannot *not* be itself. It's bearing a certain life-form is *definitive* of the kind of being it is. So, if it wants to carry on being itself, it had better maintain the processes which constitute its form (cf. Boyle and Lavin 2010: 184f.).

This normative, governing relation between a life-form and its bearers connects to the interdependence, and corresponding teleology, between the generic processes constituting a life-form. It's because these processes mutually depend on each other, that to uphold them becomes the *norm*, the thing to do, for individual members of the corresponding life-form. Otherwise these members, as bearers of this life-form, would cease to exist. Since this purposive interdependence is *essential* to life-forms, this means that the normativity inscribed in intentional (purposive) descriptions of life-processes is not an *additional* but a *constitutive* feature of life-forms. Furthermore, since, on M. Thompson's view, any judgement of a thing as *alive* presupposes the wider context a life-form, it also means that the recognition of anything as *living* already brings with it a normative standard. In M. Thompson's words 'we […] go no farther for critique than we went for

interpretation' (2008: 81). Put differently, judgements of natural goodness do not fallaciously derive *ought* from *is*. They give expression to an evaluative dimension which is inherent to the nature of life itself (Foot 2001: 36ff.; cf. Boyle and Lavin 2010: 184f., Haase 2018: 93f.). In sum, life-forms illuminate the notion of a self-enabling network of generic processes and the normativity internal to the teleological-explanatory form of processes described in the intentional order. Indeed, if we follow M. Thompson, then, it seems, our notion of a productive process and his notion of a life-process simply coincide.

2.7.2 Remaining Challenges

However, challenges remain. First, as Lewens (2012, 2020) argues, the relationship between the 'species-natures' of life-forms and the biological facts is underdetermined. It's simply unclear what makes NHJs *true*, or whether they even can be true (or false). Second, the conception of a *self-enabling system* which life-forms embody is too abstract to explain the applicability of *thick* teleological explanations to life-processes. As noted earlier, to explain how a productive process can be its own *source*, we must understand how the generic processes of the self-enabling network to which it belongs *depend on* and *precondition* each other. M. Thompson does not explicate the 'inner causal structure' (2008: 79) of dependence between life-form-processes beyond saying that 'the teleological connective expresses the concept that is converse to this conception of dependence' (ibid.). The teleological connective he characterises in terms of 'in order to' (e.g. as in our pine-example), a characterisation he (ibid.: 77) attributes to Foot (2001: 30-32). However, we saw earlier that 'in order to' does not capture the idea of *thick* teleological explanation. When we say 'The pine opens its cones *in order to* release the seeds', why is it that the cones' opening is represented as *coming from the pine*? Sure, the sentence expresses a *dependence* between the cones' opening and seed release. But how does this dependence differ from the one expressed in 'The wind is blowing *in order to* pick up the leaves'? Why does only the first sentence describe a productive, *self*-moving process? To answer this question, we said earlier, we must understand the self-enabling form of the subject as a network of interdependent generic processes. However, if the interdependence of these processes is itself understood via 'in order to', we encounter the same problem as before: we are employing a conception of dependence which is *too thin* to explain the concept of a process' accounting for its *own* realisation.

Finally, life-forms do not explain, but presuppose, the notion of a *concrete individual organism*. M. Thompson's conception of this notion derives entirely from the concept of a life-form, i.e. an individual organism is 'anything that bears a life-form'. But then presumably any pine cone is an individual organism, and so is the heart of my neighbour's cat. To rule out such oddities, we must somehow distinguish between a concrete individual, i.e. *this* pine, *my neighbour's cat*, and its parts. To mark this difference, we require a further distinction, namely that between a concrete individual and what we called its physical environment, and we established earlier that life-*forms* cannot draw this distinction. However – for reasons elaborated in Ch. 3 – the notion of a concrete individual organism, as distinct from its parts and its environment, is essential to understanding agency.

In raising these issues, I am not seeking to reject the notion of a life-form. Indeed, these points leave its main contribution, i.e. the formal representation of the 'special turn' (M. Thompson 2008: 27) which thought undergoes when it concerns *life*, untouched. Rather, I wish to flag the challenges facing any account of productive processes, regarding (i) its factual basis (ii) the notion of a self-enabling system of generic processes and (iii) the notion of individuality. With these challenges in mind, let us return to the enactive arena and the relationship between life and agency.

3 The Continuity between Life and Action

3.1 Connecting the Dots

At the end of Ch. 1, we raised the following questions.

(A) Is life necessary for agency?
(B) Is life necessary for cognition?

How do these question relate to our discussion of productive processes in Ch. 2? To answer this question, we must first extend the notion of a productive process to *systems*. Building on condition (ii/b) of our definition in Ch. 2, Sect. 2.4.2 let us say:

> A system S is productive *iff* S's form is a self-enabling network of generic processes

Thus understood, not everything a particular system 'does' must be a productive process for the system to be productive. It is so only if (ii/a) and (ii/c) are satisfied in addition to (ii/b). Like agency, productivity, i.e. being productive, should be understood as a *capacity* which needn't always be exercised. In that, our understanding of a productive system derives from our understanding of productive processes, as tied to thick teleological explanations. Being productive means: having the capacity of engaging in productive processes; a capacity which, in turn, is based on the system's bearing a certain kind of self-enabling form.

To connect our topics from Ch. 1 and 2, I will assume the following: *actions are productive processes*; or similarly: *systems capable of agency, i.e. agents, are productive systems*. This assumption should be uncontroversial. For an action is, paradigmatically, a process admitting of a thick teleological explanation. It's a process, e.g. going upstairs, carried out in view of realising another, larger process, e.g. getting the camera, in terms of which it can be explained, both as the larger process' part and, more importantly, as its *cause* or *source*. An action, like a productive process, is a kind of self-movement. In assuming that actions are productive processes, we are *not* assuming (or denying) that every productive process is an action. We are saying that any action is, among other things, a productive process.

Given our discussion in Ch. 2, any satisfactory account of a productive process should satisfy two explanatory constraints:

(C1) Explain of the notion of a *self-enabling system of generic processes*.
(C2) Provide a normative standard which is (a) *general* and (b) *governs* the subject's activity.

If an action is a productive process, and hence an agent is a productive system, these constraints also apply to an account of agency. Again, this does not mean that, in meeting these constraints, we have explained what an agent is. It means that, to explain the notion of agency, we must, among other things, satisfy (C1) and (C2).

These constraints articulate minimal explanatory demands for any conception of agency.

The first part (Sect. 3.2-3.3) of this chapter examines whether the enactive conception of agency can meet these constraints. We shall see that it can, yielding an enactive-style conception of a productive process and system, respectively. The second part (Sect. 3.4. onwards) then investigates whether we can make sense of this conception apart from a productive system's being a *living* system. I will suggest that we cannot and that therefore, since any agent is a productive system, life is necessary for agency.

3.2 Grounding Normativity

Does the enactive conception of agency meet (C2)? Before addressing this question directly, let us approach it through considering a certain tension between the enactive understanding of individuality and normativity.

3.2.1 Autonomy, Individuality and Normativity

As explained, Autonomy seeks to account for a system's *individuality* by providing a clear, observer-independent distinction between a system and its environment through the notion of organisational closure. However, autonomy also serves as the bedrock for *normativity*, providing a fundamental reference-point for (observer-independent) evaluations of the system's states and activities.

However, Autonomy, it seems, cannot account for *both* normativity and individuality, at least not if individuality is understood in a certain way. Individuality is commonly understood as *numerical identity*, which, in turn, is usually explained via Leibniz's Law that x is identical with y *iff* everything true of x is true of y. As Noonan and Curtis (2022: Sect. 2) note, Leibniz's Law underwrites our grasp of what makes an individual *distinct*, i.e. what distinguishes it, as *this particular thing*, from what it's not. Hence, if Autonomy accounts for numerical identity, then, by Leibniz's Law, it would have to distinguish a system from everything it is not (including other, similar systems). Suppose that Autonomy, by specifying a system's organisation, satisfies this demand. Then, Autonomy cannot *also* provide a normative standard. For we said earlier that a standard must leave room for the possibility of an individual's *failing* to meet it. But a numerical individual cannot fail to be identical to itself. Hence, if Autonomy defines its numerical identity, it will be impossible for the individual not to uphold its Autonomy. Absolutely anything the system does will satisfy its governing 'standard', i.e. its own identity. A puffin flying against a cliff or a bacterium immersed in a poisonous substance would be equally 'viable' with respect to this particular individual's 'standard'. No comparison could be made to what other puffins or bacteria are doing since, by hypothesis, these are different Autonomous individuals, subject to different 'standards'.

But have I not overlooked adaptivity? Indeed, the preceding point resembles Di Paolo's that autopoiesis alone allows only black-or-white evaluative judgements: either the system's integrity is intact, or it is not. From that perspective,

flying against a cliff versus catching fish are equally 'viable' behaviours, provided the puffin survives (i.e. conserves its organisation). However, although adaptivity changes the richness of the system's normative domain (its world, in Varela's words), allowing for graded, projective evaluations, it leaves its fundamental reference-point untouched: the maintenance of the system's autopoietic (or, more generally, Autonomous) organisation (Di Paolo 2005: 438-442). That is, 'The boundary of viability' cited in the definition of adaptivity remains determined by a system's Autonomy. Furthermore, 'adaptivity works mainly on tendencies of states' (ibid.: 442.), which 'relate it as a whole to the potential loss of its viability' (ibid.: 438). That is, adaptivity functions through the system's recognition of, and regulative responses to, current (internal or external) situations as having the *tendency* to improve or impair its chances of maintaining Autonomy. As Di Paolo et. al. write elsewhere, adaptivity presupposes the system's recognition of the *non-actual, potential* consequences of its current situation, which constitute its 'virtual field' (2018: 33). Thus, adaptive, graded evaluations remain dependent on Autonomy for their *sense*. Adaptivity does not introduce a *new* standard for an Autonomous system, it enriches the space of possibilities through which a system can fail or succeed in meeting a standard it already has. Adaptivity, then, doesn't solve the problem that, *if* Autonomy accounts for individuality, understood as numerical identity, it cannot also account for normativity. Indeed, if Autonomy accounts for individuality, the very idea of a *tendency*, central to adaptivity, is suspect (see below).

The problem is not only that, if Autonomy accounts for individuality, it justifies only all-or-nothing evaluations. It is that, if autonomy accounts for individuality, we cannot so much as *make sense* of evaluative judgements relative to the system's Autonomy. To ground normative evaluations, the relevant standard must be *more general* than the particular individual itself. Consequently, if Autonomy is to provide a normative standard, it cannot also account for individuality, at least not in the sense of numerical identity. For a standard is *general* precisely because, in principle, it can govern many individuals. But something which applies to many individuals will, by Leibniz's Law, not account for numerical identity.

3.2.2 Autonomy and Generality

The foregoing argument is perhaps somewhat terse. Nevertheless, its conclusion articulates a point already implicit in the concept of Autonomy, namely that the individuality which Autonomy accounts for is not numerical identity. *The Autonomous system* which Varela et. al. discuss is more an individual like *the bobcat* or *the scots pine* are individuals in 'the bobcat breeds in spring' (M. Thompson 2008: 70) or 'the Scots pine has ovoid cones'. Just as a life-form, an Autonomous organisation defines a *generic* kind of system, determining a set of individual systems sharing this organisation (Maturana and Varela 1980: xix-xx; E. Thompson 2007: 44). As Di Paolo writes, expanding on autopoiesis, a system's organisation defines a 'class identity' (2018: 79), where 'different instances in this class can be said to belong together, regardless of how much they differ in terms of how their organization is actually instantiated' (ibid: 78). Whilst instances of a

given Autonomous organisation are numerical individuals, insofar as they instantiate the same Autonomous organisation, they are *as one*, sharing exactly the same properties: 'any predicate based on a logical analysis of the properties of a given organization, will *ipso facto* apply to all instances of a class' (ibid. Original emphasis). If we understand Autonomy thus generically, then, similarly to M. Thompson's (2008: 76f.) life-form-dependent conception of an individual organism, being a *particular* Autonomous system simply means: being a numerical instance (something you can count) of a certain *kind* of Autonomous organisation.

Autonomy, understood generically, provides a standard which satisfies (C2). A generic Autonomous organisation is general enough to govern the operations of different systems (its instances). Incidentally, it will be possible for any particular instance to engage in a process which conflicts with and thus *fails* to uphold the Autonomous organisation which defines its generic form. Moreover, any particular system, by being such an *instance*, will also be *governed* by the generic Autonomous form it exemplifies. Although it may differ from other instances in certain ways, e.g. its size, age or spatial position, its Autonomous form remains an integral part of its identity (just a puffling and its mother, although different, are both essentially *puffins*). Thus, as with bearers of a life-form, an individual Autonomous system cannot *not* uphold its Autonomy without contradicting its own nature.

On this picture, 'the domain of possible interactions' in condition (iii) of Autonomy (see Ch. 1), represents a kind of norm-catalogue for any particular Autonomous system. If *consuming such-and-such materials* is a way of maintaining its Autonomous form, then the system, e.g. a bacterium, *because* it instantiates this form, may be evaluated relative to it. If it does not consume such-and-such materials, e.g. sucrose, in a situation where, other things being equal, consuming sucrose would be *the thing to do* for members of its Autonomous kind, we can evaluate this situation as a *failure* on part of the system.

Thinking of Autonomy in generic terms also bolsters the adaptive notion of a tendency. If Autonomy only captured the numerical identity of a system *here and now*, it would be unclear how, on the basis of its Autonomy, we can say that, e.g. the system's current activity *tends towards* a breakdown of viability. Absent any standard of comparison apart from the current organisation of the system itself, on what basis could we say that it is about to break down?[20]

Of course, an observer might say of a particular system 'If this puffin keeps flying like this, it will crash and loose its wings', i.e. its current activity *tends* to compromise its viability. But if Autonomy defines numerical identity, this judgement would not be original to the system. By identify the individual as a *puffin* already likens it to a more general kind, from whose perspective the judgement makes sense. But, by hypothesis, we can do no such thing. All we can consider is the system's organisation, considered by itself. *On that basis*, who's to say that 'loosing' its 'wings' tends to compromise the system's viability? Perhaps what we call 'wings' are really mating-outgrowths to be stripped off in a selfless act of cliff-crashing. Is, then, the system's behaviour viable after all (cf. M. Thompson 2008: 54ff.)?

[20] The situation is like encountering an unknown life-form. Our evaluative judgements concerning its operations cannot get a grip because, initially, there is no 'wider context' (M. Thompson 2004).

These points recall M. Thompson's insight that, in describing anything as a life-process, we presuppose a wider context (a life-form, in his terms), where 'wider' minimally means: *more general* than what we can see here and now. They also recall Boyle and Lavin's point that any progressive judgement, *S is doing A*, must presuppose a conception of the general form of the subject S. To describe a current situation as *tending towards* an outcome, which may be seen as good or bad for the relevant system, presupposes the kind of generic judgement these authors discuss (NHJs or FC-judgements). Moreover, if, as adaptivity demands, these judgements are to be based on the system's *own* organisation, as opposed to an observer's interpretation of it, then the system itself must embody a general kind of organisation. Autonomy, understood generically, can serve this job. Going forward, let us assume that Autonomy provides a standard which is general and governs the processes of an individual system, thereby satisfying (C2).

3.3 Grounding Productivity

Can the enactive conception of agency meet (C1), i.e. explain the notion of a self-enabling system of generic processes? Given the previous section, the answer might seem obvious. After all, an Autonomous system is a self-enabling system of generic processes par excellence. The notion of organisational closure, especially, supplies a tidy tool for determining the systems of interest (at a given level of observation) (Di Paolo et. Al. 2017: 112-114; 128). Moreover, given its footing in autopoietic theory and subsequent developments, organisational closure is firmly rooted in the biological and cognitive sciences (E. Thompson 2007, Moreno and Mossio 2015). Thus, *prima facie* Autonomy cannot be criticised as being detached from the biological facts as the notion of a life-form can.

On a closer look, however, it is less obvious whether Autonomy adequately meets (C1). Recall that the notion of a self-enabling system of generic processes was introduced in order to account for the thick teleological explanations characteristic of a productive process. More specifically, it was introduced as a way of explaining how a process can be its own source by being its own *precondition*. The notion of precondition, in turn, was tied to the idea of a cause which accounts for the occurrence of a process as being *non-arbitrary*, i.e. as upholding, as Boyle and Lavin put it, 'the very conditions that make its own coming to be *no accident*' (2010: 182. My emphasis). The notion of a self-enabling system of generic processes provides an abstract conception of how a particular process can precondition itself, i.e. by instantiating a generic process belonging to a network on which the process itself depends. However, the notion of such a network presupposes the concept of *self-enablement*, understood through the non-accidental, interdependent, and circular relationship between its component processes.

Unless we understand precisely how an Autonomous system enables itself non-accidentally, we do not fully understand how thick teleological explanations apply to the processes which constitute such a system as an organisationally closed unity. Consequently, we do not understand how particular instances of such a system can engage in processes which account for their *own* realisation, which

precondition *themselves*. To satisfy (C2) such as to explain the notion of thick teleological explanation, we must substantiate the idea of self-enablement.

3.3.1 Interactional Asymmetry Revisited

We might approach self-enablement via interactional asymmetry. As explained earlier, interactional asymmetry captures the idea of a system's being the source of its own activity. It thus promises to explain how a process admits of a thick teleological explanation, i.e. a kind of explanation which characterises the occurrence of the process as something which the system itself brings about. However, the enactive conception of interactional asymmetry does not really explain *why* a process can be said to originate from the system or cause (in the sense of non-accidentally enabling) itself. Let me explain.

Interactional asymmetry was defined as a system's capacity (in certain circumstances) to modulate its coupling relation to the environment. A resurrection plant which curls up into a ball asymmetrically improves its conditions of survival, not by directly acting on environmental processes themselves, but by changing the way environmental processes impinge on it. But on what basis can we say that the curling up is a process which originates in the plant, rather than its environment? Sure, the coupling relation is being modulated, but why is this modulation *coming from the plant*, why is it the *source* of what's happening? We can clarify the issue here by considering Di Paolo et. al.'s formalisation of interactional asymmetry:

$$\frac{d\mathbf{x}}{dt} = S\left(\mathbf{x}, \mathbf{p_Q}(\mathbf{e})\right)$$

$$\frac{d\mathbf{e}}{dt} = E\left(\mathbf{e}, \mathbf{p_R}(\mathbf{x})\right)$$

$$\{\mathbf{r}\} \subset \mathrm{QUR}$$

$$\Delta\, \mathbf{r} = H_T(\mathbf{X})$$

S and *E* denote functions, each governed by a set of parameters $\mathbf{p_Q}$ and $\mathbf{p_R}$, which respectively belong to sets of conditions and constraints Q (representing the agent) and R (representing the environment) and which act on **e** and **x**, representing the system and the environment respectively (understood as vectors of variables). The system-environment coupling consists in the influence of *E* on *S*, and vice versa, in relation to parameter changes in $\mathbf{p_Q}$ or $\mathbf{p_R}$. According to Di Paolo et. al. '[a]symmetric modulation of this coupling is described by the system's influence on a subset $\{\mathbf{r}\}$ of these constraints [i.e. QUR] […], which the system modifies according to the function H_T that depends on its own states **x**, and which is active during a particular time interval *T*.' (2017: 118). The key phrase, for our purposes, is 'according to the function H_T that depends on its own states **x**'. This phrase shows that the formalisation *assumes* that the modulation of the system-environment coupling ($\Delta\, \mathbf{r}$) is driven by the system ('depends on its own states'). This assumption is fine, of course, but it does not explain, but presupposes, that we know what it means to say that the modulation depends on the system's states. The notion

of interactional asymmetry, as it stands, does not seem to solve that issue. Rather, it transfers it to a meta-level, shifting the question in virtue of what the system's *direct* effect on the environment comes from the system to the question in virtue of what the system's *indirect* effect on the environment, via modulation of its coupling, comes from the system.

The issue is reinforced by Di Paolo et. al.'s (2017: 120) acknowledgement that asymmetrical changes sometimes originate from the environment, including other agents. Again, this does not compromise the definition since interactional asymmetry is defined as a system's *capacity* which need not be realised constantly. Nevertheless, we do not understand this capacity unless we understand what it means to say that a modulation *comes from the system*. One might distinguish the relevant modulations by means of the other two criteria of agency, notably normativity. To take the authors' (ibid.: 124) example, the spasms of a Parkinson's patient, despite initiating an asymmetrical modulation, cannot be considered an action because they do not satisfy a norm or goal of the patient. But then again, there are asymmetrical couplings of self-individuating systems in which intrinsic norms are satisfied, but which nonetheless do not originate from the system, e.g. a bacterium put in a sucrose solution by an experimenter. To fully explain self-enablement, it seems, interactional asymmetry is not enough.

3.3.2 A Precarious Promise

A more promising candidate to account for self-enablement is precariousness. This concept was introduced to exclude trivial cases of organisational closure, e.g. snowflakes and crystals, and, more importantly, to explain why, to continue its existence, an Autonomous system must adaptively uphold its organisation. Because a precarious system's processes deteriorate in the absence of the enabling conditions provided by the organisationally closed network, the system is constantly preoccupied with maintaining its own organisation, e.g. repairing and attuning relations between processes; regenerating, recycling, and discarding components etc. As mentioned before, a precarious system is 'literally self-enabling' (Di Paolo and Thompson 2014: 72): it actively sustains the conditions of its own existence (Di Paolo 2009: 16).

Precariousness, together with Autonomy (organisational closure), gives a clear sense to the idea of a process' being its own precondition, which accounts, non-accidentally, for its *own* realisation. As such, it substantiates the notion of a productive process' being its own source, which justifies the applicability of thick teleological explanations. A precarious process could not occur unless it contributed to upholding a closed system of processes on whose existence it itself depends. In that sense, a precarious process owes its occurrence to its own activity (or more precisely, to the end-result which marks its completion, which in turn enables another process): it preconditions itself. Furthermore, this self-preconditioning is non-accidental because the very network of processes to whose existence a precarious process contributes, and on which it depends, is not a happenstance or (merely) the result of favourable environmental conditions. The system's existence is something which this process, in union with others, perpetually *makes happen*. In that sense, a precarious process contributes to 'the

very conditions that make its own coming to be no accident' (Boyle and Lavin 2010: 182). Thus, precariousness, combined with Autonomy (understood generically), satisfies (C1).

3.3.3 Productive Processes, Enactive Style

Given the previous sections, we might propose the following enactive grounding of a productive process:

> A process, *S is doing A*, is productive *iff*
>
> (i) *S is doing A* admits of a thick teleological explanation in terms of some other process, *S is doing B*. [where, as a limiting case, ◇ (*S is doing A = S is doing B*)]
> (ii) Where *S is doing A* admits of a thick teleological explanation in terms of *S is doing B iff*
> (a) *S is doing A* is a lesser or identical stage of *S is doing B*
> (b) S's form is an Autonomous (organisationally closed) system of generic precarious processes
> (c) *S is doing B* is an instance of a generic process in this Autonomous system

The notion of an Autonomous system of generic precarious processes in (ii/b) replaces the previous idea of a self-enabling network. The generic form of this system is explained through organisational closure, whilst the aspect of self-enablement is captured by precariousness. There is, however, a conceptual interdependence between closure and precariousness which, to some extent, preserves the Anscombian thought that the description of a process as productive – or an action (Anscombe), or a life-process (M. Thompson) – is *as such* the description of something which exemplifies a certain form or wider context 'through which' it is apprehended. For our understanding of a precarious process is inextricably tied to our understanding of its place within the generic Autonomous form of its subject. In Anscombe's terms, our description of something as a precarious process could not occur *prior* to our understanding of how it depends and contributes to the existence of this form. In that sense, the dependence of a precarious process on the enabling relations constituting the Autonomous network to which it belongs is not only causal-effectual, but also formal or conceptual. Since, on our proposal, a productive process is (among other things) a precarious process, our characterisation embodies the Anscombian thought whilst offering a naturalised understanding of it through the enactive idea of Autonomy. Indeed, given that, following Anscombe, the form invoked in our proposal is *teleological* (end-directed) it also – in line with E. Thompson (2007: Ch. 6) – preserves the Kantian thought that our apprehension of self-organising beings as ends in themselves is the 'ground for the cognition of the systematic unity of the form and the combination of all of the manifold that is contained in the given material for someone who judges it' (Kant 2000: 245).

But is not the *form* which Anscombe, M. Thompson and Kant talk about purely intellectual (M. Thompson 2008: 79) or transcendental (Rödl 2012)? If so, can it be 'grounded' in an empirically based notion like Autonomy? I will not attempt to answer these questions here. Let me just emphasise that, in saying that our proposal naturalises 'the Anscombian thought', we are not claiming that it *reductively explains* (thick) teleological explanations or the form of intentional descriptions in non-teleological, e.g. mechanistic, terms, as some authors (e.g. Ward and Villalobos 2016a, 2016b) suggest is required of naturalistic explanations. On the contrary, understanding the teleological form of purposive, intentional descriptions seems to be essential to our grasp of Autonomy. In this, our proposal resonates with Weber and Varela's realist turn towards teleology as revealing 'a dynamic pattern of the activity proper to life that would otherwise be missed' (E. Thompson 2007: 146) and with enactivism's self-understanding as a non-reductive naturalism (Di Paolo et. al. 2010; Di Paolo 2018).

According to our proposal, an intentional description of a process, e.g. Beth is going upstairs (in order to A), the dog is holding a bone, the wind is picking up the leaves, truly describes a *productive* process, i.e. truly ascribes a *thick* teleological explanation, just in case the process in question is (part of) a process which instantiates a generic precarious process within the Autonomous organisation of generic precarious processes which constitutes the subject's form (it's quite a mouthful, I admit). The dog's holding a bone, for example, is a productive process by being a stage in a larger process, e.g. eating, which instantiates a generic process forming part of the Autonomous system(s) which constitute(s) the generic form *dog* (the '(s)' here will become clearer later). *Eating*, itself divisible into many smaller processes, is a characteristic description of a generic *precarious* process, i.e. it is a process which contributes to and depends on a network of processes, e.g. digestion, excretion, etc. without which it would cease to be possible. Accordingly, 'the dog is holding a bone *because* it is eating it' expresses a thick teleological explanation: it is the description of a process which originates from the subject, in a purposeful, non-accidental manner. The wind's picking up the leaves, in contrast, is *not* a productive process. Sure, *picking up* is a *description* of a process which, paradigmatically, does originate from the subject, e.g. as when a dog picks up a bone. But *the wind* is not an Autonomous system in which the process of picking up leaves plays a precarious role. As this suggests, our characterisation can identify animalising descriptions: they are intentional, purpose-oriented descriptions of a process whose subject is represented *as if* it was an Autonomous (organisationally closed) and precarious system, when really it is not.

3.4 Productive Autonomy, Precariousness and Materiality

Let us return to our guiding questions, focussing particularly on (A). First of all, let us transfer the notion of being productive to systems:

> A system S is productive *iff* S's form is an Autonomous (organisationally closed) system of generic precarious processes

How does this concept of a productive system bear on the question whether life is necessary for agency? The connection is simple. If a productive system is, necessarily, a *living* system, then, given our assumption that every agent is a productive system, life will be necessary for agency. Since life is adaptive autopoiesis, what we need to investigate is the relationship between adaptive autopoiesis and productive systems. We can provisionally ignore adaptivity. Let us call the Autonomy of a productive system *productive Autonomy*. Our question, then, is the following: is a productive system necessarily an autopoietic system, i.e. is autopoiesis necessary for productive Autonomy?

As explained in Ch. 1, what distinguishes autopoiesis from other forms of Autonomy is (a) a topological (material) boundary, e.g. a membrane and (b) that the system's component processes are processes of production (chemical, molecular reactions). To avoid confusion, I shall use 'productive process' strictly as just defined and 'process of production' for the component processes of an autopoietic system. Given (a) and (b), we can reformulate our question: does productive Autonomy necessarily involve the construction of a topological boundary and that the component processes of a productive system are processes of production? Let us focus on the latter part of this question. Since the component processes of a productive system are *precarious*, our immediate concern is the relationship between precarious processes and processes of production. (How) are these notions connected? Let us prepare our reply via an insightful detour through some ancient territory.

3.4.1 Charles on Aristotle on Inextricably Enmattered Forms

In *The Undivided Self*, Charles argues that, for Aristotle, the forms of natural objects are 'inextricably enmattered, inseparable in definition from material features or activities' (2021: 53). Leaving aside the interpretation of Aristotle, what interests us is Charles' notion of *enmattered form.*

According to Charles, a property P exhibits what he calls an 'S-structure' just in case

> (i) being P is defined as a distinctive way of being Q and
> (ii) being P is not defined by decomposition into two or more terms, one of which refers to a more general well-defined property or nature (Q*) shared by other properties (2021: 51f.)

An S-structure is a generalisation of (Charles' reading of) Aristotle's understanding of *snubness* (hence the 'S'). Aristotle writes

> "concave" signifies (*sēmainei*) the same thing when used commonly of the snub and of the bandy (a type of leg). However, when added to the nose and the leg it can signify different things. In the former case it signifies the snub, in the latter the bandy; and, in this case, there is no difference between saying snub nose and concave nose […] there is nothing strange if the snub nose is the nose having nasal-concavity. (*Sophistici Elenchi* 181b34-182a6, as quoted in Charles 2021: 47).

Initially, this passage suggests that both snub noses and bandy legs share a common feature: concavity. However, Aristotle insists that 'concavity', when attached to the nose, means something different as when ascribed to legs. In the former case, it signifies a distinctively nasal kind of concavity, nasal-concavity, which cannot be ascribed to legs. This kind of concavity, Charles explains, 'cannot be defined as a combination of concavity (defined independently of the nose) and the nose.' (2021: 47). That is, *concavity* is not an 'independent component' (ibid.: 48) in the compound 'snub nose'. Unlike 'being spherical' the concavity of snubness is not a well-defined property shared by different compounds, as bronze spheres and golden spheres are both spherical in the same way (ibid.). Since, when applied to snub noses, 'concavity' signifies nasal-concavity, it cannot be used to describe e.g. bandy legs. Snubness thus exhibits an S-structure. It is (i) a distinctive, i.e. inextricably nasal kind of being concave and (ii) there is no one well-defined common property, concavity, referred to in the definition of snubness and bandiness (or any other property) into which snubness could be decomposed (ibid.: 49).

A property exhibiting an S-structure, i.e. which is like snubness, is an *inextricably enmattered form.* The reference to its matter is part of the *definition* of the form, e.g. as being nasal is part of the definition of the kind of concavity which snubness is. Is productive Autonomy an inextricably enmattered form of Autonomy? If so, then

(i)* being productively Autonomous is defined as a distinctive way of being Autonomous

(i)* being productively Autonomous is not defined by decomposition into two or more terms, one of which refers to a more general well-defined property, Autonomy*, shared by other properties

To show that productive autonomy is inextricably enmattered, there must be a *distinctive, matter-involving way* of being Autonomous, which other, non-productive kinds of Autonomy* lack and which is part of the definition of productive Autonomy (the asterisk indicates that 'Autonomy', when applied to *non*-productive systems, signifies something else as when applied to productive systems).

Autonomy, as it figures in our definition of a productive system, has two aspects: organisational closure and precariousness. Organisational closure, as we have seen, is a property which is independent of the nature of the processes it characterises (E. Thompson 2011: 215). Accordingly, we are unlikely to find a matter-involving aspect in the closure-aspect of productive Autonomy. But what about precariousness? If precariousness is a property which is inextricably enmattered, then, given its central place in our understanding of productive systems, the Autonomy of such systems, i.e. productive Autonomy, will also be inextricably enmattered. Can precariousness be understood independently of any reference to matter?

3.4.2 Hans Jonas on the Fragility of Metabolism

This question might be approached differently. For Charles' Aristotle, the forms of natural objects had to be enmattered because, otherwise, they couldn't play the causal roles required of them (Charles 2021: 68ff.). Similarly, we might consider the causal-explanatory role which precariousness plays in relation to self-movement, i.e. a process' being its own source. However, the topic of causation is too complex to be properly examined here (see e.g. Sosa and Tooley 1993). Instead, we will scrutinise the inspiration of the concept of precariousness: Hans Jonas' analysis of metabolism (Weber and Varela 2002, Di Paolo 2005, 2009, E. Thompson 2007: Ch. 6, Froese 2017).

According to Jonas, 'metabolism can well serve as the defining property of life: all living things have it, no nonliving thing has it' (1996: 88). Metabolism denotes the organism's maintenance of its living form throughout the flow of matter. For Jonas, however, metabolism is not merely a feature of organisms. It is the defining ontological principle of their *mode of being*. It pervades all other characteristics of life, e.g. reproduction and growth, and 'distinguishes the living during *any* stretch of its existence […] from the non-living' (2016: 67. Original emphasis). Metabolism is a mode of existence which essentially relies on the constant exchange of materials. Consequently, life is by its very nature inherently fragile, ever-lacking, and related to mortality (1996: 88f.). For an organism cannot completely possess the matter of which it consists, which traverses its organic form. Being reliant on a constant supply of it to uphold its metabolism, an organism's life is marked by the incessant pursuit of matter. Therein lies, for Jonas, the fundamental distinction between the mode of being of living and non-living things. A non-living entity, e.g. a material particle, 'is simply what it is, immediately identical with itself without [the] need to maintain that self-identity as an act of its existence' (Jonas 2001: 81). But a metabolising, i.e. living, system can only exist '*by* not remaining the same matter' (ibid.: 76. Original emphasis). Whilst dead particles exist unconditionally, so to speak, a living being *only* exists through its own activity; or indeed, its existence *is* its own activity: 'organisms are entities whose being *is* their own doing […] *doing what they do is their being itself*' (Jonas 1996: 88. My emphasis).

Jonas' position shares Anscombe's Aristotelian spirit. It's not only, Jonas thinks, that living things have a property, i.e. metabolism, which dead things lack. Rather, the description of a thing as a metabolising entity is the description of something which, ontologically speaking, resides in a different sphere of being from non-living things. Living things have a fundamentally different *kind of identity* from the non-living, which differs not only in its properties, but *in the way in which properties can be ascribed to it* (cf. Stoutland 2011: 31f.). This also emerges from Jonas' reflections on the intrinsically teleological nature of life, whose enactive appropriation has sparked much controversy.[21] Since a metabolising system must constantly gather resources from its surroundings, its activity aims *beyond* itself, towards its environment (Jonas 2001: 84ff.; cf. Barbaras 2010, Coyne 2021: 57). It is oriented towards what is currently lacking, but required for the maintenance of metabolism. Life is, in Jonas' words, *self-transcendent* (2001: 86f.). Jonas

[21] See e.g. Ward and Villalobos (2016a, 2016b), De Jesus (2016), Kee (2021), Hverven and Netland (2021), Prokop (forthcoming).

understands this transcendence not only spatially, as movement towards external objects, but also *temporally*. Since an organism must constantly be active to uphold its metabolic mode of being, its existence is shaped by what it's *going to be*, rather than by what it was. An organism's mode of being is inherently forward-looking, it transcends into the future. This temporal transcendence, Jonas writes, 'is the root of the teleological or finalistic nature of life' (ibid.: 86), a finalism which is 'in the first place a dynamic character of a certain mode of existence […], and only in the second place a fact of structure or physical organization […]. (ibid.). Unsurprisingly, Jonas calls this conception of teleology an 'Aristotelian reminder' (ibid.): it parallels the Anscombian thought that goal-directedness is primarily a mode of existence, an order of being, permeating a form of judgement and description. Without grasping this dynamic mode of being which defines the organic identity of living things, Jonas thinks, 'eye would not be recognized as eye, feeler not as feeler, organism not as organism' (ibid.: 87) and scientists 'would altogether miss the existence of life around them' (ibid.: 91). Compare this to M. Thompson's suggestion that, without grasping the teleological, generic form of judgement characteristic of life-forms, an intellect will lose 'the capacity to experience things as alive. It can no longer look to a "wider context"' (2008: 77).

Jonas' analysis of metabolism emphasises two interrelated ideas. First, because of its fragile relationship to matter, life is fundamentally *concerned* with the continuation of its being. This self-concern is, for Jonas, the driving force of life's ongoing, self-preserving activity in search for vital materials required to carry on its metabolic being (2001: 84f; 2016: 2-5; 58f.). Second, since life's mode of existence consists in a (n)ever-ending *exchange* of materials, the identity of organic forms is also, somewhat paradoxically, independent from its material constituents. An organism continues to be itself even though its material components at one moment might be entirely replaced at another (Jonas 1968). In Jonas' words 'the organic form stands in a dialectical relation of *needful freedom* to matter' (2001: 80. Original emphasis).

3.4.3 Is Productive Autonomy Inextricably Enmattered?

Jonas' reflections on life's fragility and concernedness mainly inspired enactivism's normative turn (Barrett 2017: 437). Since the concept of precariousness, especially, derives from Jonas, it seems reasonable to extend the implications of Jonas' reflections to our question: can precariousness be understood independently of any reference to matter? For Jonas, the answer is 'no'. Given that the precariousness of metabolising (i.e. living) systems is grounded it their needful relationship to materiality, which is integral to their dynamic, teleological mode of being, it cannot be understood without reference to matter. Living forms, insofar as they are precarious, are inextricably enmattered forms. This does not only mean organic materiality must be recognised or added when defining a living form. It means that we cannot comprehend the nature of living forms, their precarious mode of existence, apart from an understanding of their dependence on matter (cf. Godfrey-Smith 2016). As Jonas acknowledges, the idea of forms persisting throughout the change of their material realisation is well-known in physics, e.g. in the mathematical description of waves (2016: 42ff.; 2001: 177). Yet, precisely because

such forms can be *defined* independently of matter, as abstract objects of mathematical contemplation, applicable, in the same way, to materially different, oscillating phenomena, that they are fundamentally distinct in kind from living forms (cf. Charles 2021: Sect. 2.4). Because they are not enmattered, their existence lacks the need of constant material renewal and active self-integration. Only in systems which are so enmattered do we encounter 'an entirely new possibility of being' (Jonas 2001: 79): life.

If precarious forms are inextricably enmattered, then, since precariousness is a necessary ingredient for productive Autonomy, the latter is likewise an inextricably enmattered form. That is:

(i)* being productively Autonomous is defined as a distinctive, i.e. precarious way of being Autonomous
(ii)* being productively Autonomous is not defined by decomposition into two or more terms, one of which refers to a more general well-defined property, Autonomy*, shared by other properties

If this is correct, productive Autonomy is a property which cannot, in the same sense, be applied to non-precarious systems. The understanding of this property is inextricably matter-involving. When we call a non-precarious system, e.g. a crystal or snowflake, 'Autonomous' we *must mean something different* as when we apply the same term to a precarious system (just as, for Charles' Aristotle, calling a line 'concave' means something different, i.e. geometrical concavity, than calling a nose 'snub', i.e. nasal-concavity). This does not exclude the possibility of there being an independent, well-defined property, e.g. organisational closure, shared by both productive (precarious) and non-productive (non-precarious) systems, e.g. crystals and living organisms. It means that productive Autonomy cannot be *defined* (fully grasped) as the *addition* of that independent property *plus* precarious materiality. Productive Autonomy is, as such, an inextricably precarious, and thus enmattered, form of Autonomy. The productive, teleological mode of being inscribed in this form cannot be understood prior to or independently of its relationship with matter. As soon as the reference to matter falls out of definition of 'productive Autonomy', we are using this term in a different sense, e.g. we are perhaps talking about organisational closure, as understood in dynamical systems theory. We are no longer talking about *productive Autonomy* proper.

3.5 Autopoiesis and Autonomy Revisited

If productive Autonomy is inextricably enmattered, how does this affect the relationship between autopoiesis and productive Autonomy? Are all productive systems necessarily autopoietic? This question connects to the issue whether autopoiesis is necessary for Autonomy, which, in turn, since sense-making is adaptive Autonomy, relates to whether *life*, i.e. adaptive, autopoietic Autonomy, is necessary for sense-making. The relationship between autopoiesis and Autonomy is discussed by Wheeler (2011) and E. Thompson (2011) in relation to E. Thompson's (2007) *Mind in Life,* henceforth '*MiL*'. By examining their discussion, we can hope to make progress on our question.

3.5.1 An Exegetical Exercise

In his commentary on *MiL*, Wheeler identifies a tension in the relationship between autopoiesis and Autonomy (Wheeler and E. Thompson do not capitalise 'autonomy'. However, they use 'autonomy' in Varela's sense, meaning Autonomy as explained in Ch. 1). On Wheeler's reading, 'the only thing that autopoiesis adds to the concept of autonomy is the dual materiality of the systemic boundary and the systemic domain of existence' (2011: 154). By 'dual materiality' Wheeler means the (a) topological (material) boundary, e.g. a membrane and (b) processes of production, i.e. biochemical processes, characteristic of autopoiesis (2011: 151), following E. Thompson's nutshell-gloss of autopoiesis as 'autonomy in the biochemical domain' (2007: 44). As Wheeler (2011: 152f.) highlights, *MiL* strongly suggests that autopoiesis is not necessary for Autonomy by acknowledging many non-autopoietic Autonomous systems, e.g. the nervous and immune system and social-territorial networks, e.g. insect colonies, animal societies, and primate bands (E. Thompson 2007: 46ff.; 65, 2011: 213, cf. Varela 1991, Di Paolo et. al. 2017, 2018). However, Wheeler notes, some passages in *MiL* whistle a different tune:

> Agency and meaning require autonomy; minimal agency and meaning require minimal autonomy. Minimal autonomy *depends* on macromolecules but requires that those macromolecules be *organized* in a particular way, namely, in the autopoietic way. (E. Thompson 2007: 160, original emphasis; as quoted in Wheeler 2011: 153).

If *minimal* Autonomy depends on the autopoietic organisation of macromolecules, then surely other forms of Autonomy do as well? As Wheeler proposes, we might resolve this tension by reading 'minimal Autonomy' as denoting minimal *biological* Autonomy, a proposal which E. Thompson presumably accepts when replying: 'when I speak of "minimal autonomy" […], I am referring to the simplest systems we know that have all the required properties for autonomy. The paradigm case is the autopoietic cell' (2011: 214). On this picture, then, the dual materiality of autopoiesis is necessary for the most basic, biological cases of Autonomy, e.g. cells and unicellular organisms, but not necessary for more complex Autonomous systems. But even if we follow this suggestion, resolving the apparent tension in *MiL*, the more important question is: why should this dual materiality, well, *matter* for understanding minimal Autonomy and agency? As Wheeler emphasises, 'it is genuinely hard to see what special substantive contribution is made to our explanation of the genesis of intentional action by recognizing that the minimal biological form of autonomy exhibits such dual materiality, given that such materiality is apparently expendable by the time that biology gets as far as a nervous system.' (2011: 154). Indeed, Wheeler continues, the concepts of Autonomy (organisational closure), structural coupling, and adaptivity seem to be *sufficient* for a naturalistic account of intentionality, sense-making and immanent purposiveness (2011: 158ff.). Why muddy this conceptual material with materiality? What's the 'special substantive contribution' of autopoiesis?

E. Thompson's reply is multi-faceted. He distinguishes two senses in which autopoiesis may be necessary for Autonomy, which we can call 'definitional' (D) and 'constitutive' (C) dependence, respectively (2011: 213-216).

(D) Every Autonomous system is, by definition, itself an autopoietic system.
(C) Every Autonomous system realises its autonomy through autopoietic constituents.

According to (D), autopoiesis is, conceptually speaking, part of Autonomy, i.e. when we call a system 'Autonomous' we *mean*, among other things, that the system is autopoietic. The qualification 'itself' in (D) is important to mark the contrast to (C). For (C) also says that every Autonomous system is autopoietic, however, not in the sense of being *itself* autopoietic, but in the sense of 'depend[ing] constitutively on autopoiesis' (E. Thompson 2011: 214), i.e. by relying on parts, e.g. cells, which are themselves autopoietic.

E. Thompson denies (D), but leans towards (C) (though he remains unsure) (2011: 213-217). Denying (D) makes sense given that, as E. Thompson notes, Varela's (1979: 55) definition of Autonomy 'leaves entirely open the processes that can be interrelated in this way' (2011: 215), where 'interrelated in this way' means, plausibly, 'organisationally closed'. Indeed, there seems to be no *conceptual* barrier against calling non-autopoietic systems, e.g. animal societies or nervous systems 'Autonomous' in Varela's sense. These systems are organisationally closed networks of processes, albeit in different domains. The fun begins with E. Thompson's reflections on (C).

First, picking on Wheeler's gloss on autopoiesis, E. Thompson insists:

> What autopoiesis adds to this specification [Varela's definition of Autonomy] is that the processes are ones that modulate molecular transformations in the chemical domain […] the crucial feature of such a biochemical instantiation of autonomy is not the dual materiality *per se* but the fact that this materiality realizes a certain organizational boundary. Put another way, the identity of an autopoietic system cannot be defined by preservation of the membrane […] it has to be defined by preservation of the network's organizational boundary […]' (2011: 215. Original emphasis).

The crucial feature(s) of autopoiesis, E. Thompson clarifies, are *not* (a) and (b), but rather what he calls 'the autopoietic organization' (2011: 212f.). As Di Paolo remarks elsewhere, 'what matters conceptually is always [the organism's] organizational boundary' (2009: 12), not the physical boundary, e.g. a membrane.

But what is *the autopoietic organisation*? What does it have that Autonomy doesn't? Why does it *matter conceptually* whether we speak of the autopoietic organisation or the Autonomous organisation? E. Thompson's stance is puzzling. Above and in other passages (2011: 212f.), he repeatedly stresses that the materiality of (a) and (b) is not the crucial defining feature of autopoiesis. Yet, he insists that the relations between the processes constituting an autopoietic system must be 'relations between processes of molecular transformation, including those that make up the boundary' (2011: 212). So, whilst the (a) physical boundary and (b) biochemical properties of the constituent processes are not *definitive* of autopoiesis, it does seem to *matter conceptually* that the autopoietic *organization* is realised through biochemical materials. But now: is or isn't materiality part of the definition of autopoiesis? Before addressing this question, let us consider another interesting aspect of E. Thompson's reply.

Faced with the question whether there could be an Autonomous system which does not depend constitutively on autopoiesis, in the sense of (C), E. Thompson, citing Barandiaran et. al. (2009), submits that 'for the system to be genuinely autonomous, it would need to (i) be an individual […] (ii) be the active source of its interactions […] and (iii) generate the norms for those interactions on the basis of its activity' (2011: 215). With a nod to Froese and Ziemke (2009), who argue that such a system must be able to generate and adaptively regulate its own sensorimotor machinery, E. Thompson doubts that such a 'genuinely autonomous' system is possible 'without something like a *metabolism*' (2011: 216. Original emphasis), which, a line later, he rewords as '*without something like an autopoietic organization for the constituents that make up the sensors, effectors, and the adaptive mechanism that links them'* (ibid. Original emphasis). These considerations fuel his sympathies towards (C). Moreover, E. Thompson proposes that Jonas' elaborations on the precarious, concerned character of organic existence, as grounded in metabolism, also support (C) (from a different angle) (2011: 216f.).

This is interesting. For our course (i) – (iii) of Barandiaran et. al. (2009) are the enactive conditions for *agency*. Hence, E. Thompson's main reason for leaning towards (C) is not the definition of Autonomy as such, but the notion of Autonomy *as it figures in the notion of agency* – 'genuine Autonomy', as he says. This also resonates with the passage about 'minimal Autonomy' quoted earlier, which occurs in the context of a discussion with Dennett concerning the roots of agency and intentionality (E. Thompson 2007: 159ff.; 2011: 214). There, E. Thompson concludes that it is the '*autopoietic organization* that is the ground from which the seeds of intentional action grow, *not macromolecules as such*.' (2007: 161. My emphasis). Thus, in sum, we seem to be left with the following claims:

(1) Not every Autonomous system is, by definition, itself an autopoietic system, i.e. (D) is false.
(2) It *matters conceptually* that the autopoietic organisation is realised through biochemical materials ('something like a metabolism').
(3) Every Autonomous system, *insofar as it is an agent*, depends constitutively on autopoiesis, i.e. (C) is true (for Autonomous *agents*).

(Although E. Thompson remains unsure about (C), I will treat him as endorsing it).

3.5.2 How Matter Matters

Let us analyse the debate, starting with (2). Here, two interpretations clash. Plausibly, we might understand 'autopoietic organisation' as 'organisational closure'. After all, to revive a point from Sect. 1.1.3, (a) and (b) seem to contribute little to understanding the autopoietic *organisation*, i.e. the way in which biochemical processes are *related*. To grasp this organisation, the concept of Autonomy, understood as organisational closure, seems enough. This interpretation is arguably what raises Wheeler's eyebrows as to the 'special substantive contribution' of the dual materiality of autopoiesis. Unfortunately, it conflicts with

E. Thompson's insistence on (2), i.e. that materiality does matter conceptually. For although one needs to study the material constitution of a system to determine whether it is organisationally closed, the *concept* of organisational closure is silent on which kinds of processes realise closure. So, again: is or isn't materiality part of the definition of autopoiesis? It is both, I think, but to see how we must again rely on ancient support.

As Charles concedes, Aristotle's account of the forms of natural objects as inextricably enmattered hides a tension: certain passages in *Metaphysics* Z.11 apparently suggest that flesh (of the nose) both is and is not part of the form (*logos*) of the snub (Charles 2021: 55f.). As Charles convincingly argues, Aristotle resolves this tension through a previously developed distinction (in *Metaphysics* Z.10) between flesh *as matter* and flesh *as a principle of form*. The former notion understands flesh as the spatially decomposable, physical parts of the nose, whereas the latter understands flesh as *a material mode of being* contained in the *form* of snubness. Whereas flesh is part of the definition of snubness *as a principle of form*, it is not so *as matter* (ibid.).

More generally, for Aristotle, the materials which manifest inextricably enmattered forms, e.g. flesh, iron, etc. figure *as principles of form* (i.e. material kinds of existence) not *as matter* (i.e. spatially divisible bits of stuff) in their definitions. As Charles (ibid: 56) notes, Aristotle's distinction matches the ontological difference between properties and their material bearers: any puffin is a puffin, but *being a puffin* is not itself a puffin. Likewise, any snub nose is a nose, but being a nose ('nasality') is not itself part of a particular snub nose: it is the material, fleshy way of being which characterises noses and which accounts for a concave nose's bearing the form *snub* (nasal-concavity) (Charles 2021: 56ff.).

Using Aristotle's distinction, we can resolve our conflict as follows: the materials of autopoiesis, i.e. the biochemical component processes, are part of the autopoietic organisation *as a principle of form*, not *as matter*. This explains E. Thompson's insistence that the physical boundary *per se* is not essential to autopoiesis. Rather, the boundary is the physical manifestation of the biochemical mode of being, understood *as a principle of form*, of the autopoietic organisation (just as a particular snub nose is the actual manifestation of the inextricably nasal concavity which snubness is). If that's correct, then (a) is not part of the *definition* of autopoiesis, whereas (b) is part of it, but only *as a principle of form*, not *as matter*. In contrast, neither (b) nor (a) are part of the definition of organisational closure, not even *as principles of form*. Whereas autopoiesis is an inextricably enmattered (biochemical) form of Autonomy, organisational closure is not.

Now, what about (3)? It seems that, here, E. Thompson must think of the autopoietic constituents on which Autonomy depends *as matter*, i.e. spatially divisible autopoietic systems, e.g. cells, not *as a principle of form*, i.e. the biochemical mode of being. Otherwise (3) would be inconsistent with (1). For suppose that the autopoietic, biochemical constituents of, say, the nervous system (a non-autopoietic Autonomous system) figure in the definition of this system *as a principle of form*. Then, given that autopoiesis is an inextricably enmattered form, the nervous system would be inextricably enmattered *in an autopoietic way*; otherwise, since autopoiesis figures, by hypothesis, as *a principle of form*, we could not properly grasp the nervous system's Autonomy. But this means (arguably) that

the nervous system would *itself* be an autopoietic system. But this conflicts with (1). If this is correct, then (3) should read, more precisely:

> (ii)* Every Autonomous system (insofar as it is an agent) depends on autopoietic constituents *as matter*.

But this is implausible for reasons which, ironically, E. Thompson himself suggests. Recall that his motivation for adopting (3) is not Varela's definition of Autonomy, but 'genuine autonomy' as required for *agency*. Let us mark this agency-relevant Autonomy (whatever it is) as 'Autonomy$_{AG}$'. As E. Thompson (2011: 216) highlights, expanding on Jonas, Autonomy$_{AG}$ cannot be understood without reference to precariousness. This matches Di Paolo et. al.'s (2017) definition of agency, which includes precariousness as a necessary condition of Autonomy. Now, we argued, following Jonas, that precarious forms are inextricably enmattered, i.e. matter must be part of their definition *as a principle of form*. Hence, since Autonomy$_{AG}$ is, by definition, a *precarious* form, whichever materiality is invoked to account for its precariousness, it must be part of the definition of Autonomy$_{AG}$ *as a principle of form.* The relevant materiality must *matter conceptually* for our understanding of Autonomy$_{AG}$.

Similarly, E. Thompson's rationale behind (3) seems to be that the biochemical, metabolic materiality of autopoietic constituents is required to grasp the precarious nature of Autonomous$_{AG}$ systems. But (3) cannot perform this job if we understand it as (3)*. For (3)* entails that the autopoietic nature of the relevant constituents does *not* matter *conceptually* for our understanding of Autonomy$_{AG}$. Just as Beth's having a snub nose *as matter* (a material object) is irrelevant for defining her human form, so the fact that the nervous system constitutively depends on autopoietic constituents *as matter* is, according to (3)*, irrelevant for *defining* the nervous system's Autonomy. Indeed, this conceptual irrelevancy is precisely what preserves the consistency between (1) and (3)*.

However, given our previous point, if we read 'Autonomous' in (3) as 'Autonomous$_{AG}$', then (3)* must be given up. Autopoiesis cannot account for the precariousness of Autonomy$_{AG}$ unless its biochemical materiality figures *as a principle of form* in the definition of Autonomy$_{AG}$. Thus, (3) should read:

> (3)** Every Autonomous$_{AG}$ system depends on the autopoiesis of its constituents *as a principle of form.*

This means, also, that the notion of Autonomy in (1) cannot be Autonomy$_{AG}$ if (1) is to be consistent with (3)**. We can reinstate the consistency by reading 'Autonomy' in (1) as 'organisational closure', which is not an enmattered property. Again, to account for precariousness, autopoietic materiality must be part of the *definition* of Autonomy$_{AG}$, i.e. every Autonomous$_{AG}$ system must, conceptually speaking, be an autopoietic system.

In sum, we reach the following position:

> (1)* Not every Autonomous (i.e. organisationally closed) system is, by definition, itself an autopoietic system.

(2) It *matters conceptually* that the autopoietic organisation is realised through biochemical materials.

(3)** Every $\text{Autonomous}_{\text{AG}}$ system depends on the autopoiesis of its constituents *as a principle of form.*

3.6 The Continuity between Life and Action

What about our question whether productive Autonomy requires autopoiesis? Well, if productive Autonomy is $\text{Autonomy}_{\text{AG}}$, then our above discussion *strongly suggests* that productive Autonomy *conceptually* depends on autopoiesis. More precisely, it strongly suggests that, among the productive processes constituting the generic Autonomous form of a productive system, there are *processes of production* in the sense of (b) of autopoiesis, whose biochemical materiality enmatters the Autonomy of the productive system *as a principle of form*, and thereby accounts for the system's essential precariousness. Given our assumption that every agent is a productive system and given that $\text{Autonomy}_{\text{AG}}$ simply is autonomy as it figures in the enactive definition of *agency*, it is highly plausible that productive Autonomy is $\text{Autonomy}_{\text{AG}}$. This conceptual convergence also comes out in the parallels between our argument that productive Autonomy must be a precariously enmattered form and E. Thompson's argument that $\text{Autonomy}_{\text{AG}}$ constitutively (and, given our interpretation, also *conceptually*) depends on the precariousness of metabolic, autopoietic constituents. Both arrive at the conclusion that productive Autonomy and $\text{Autonomy}_{\text{AG}}$, respectively, are precariously enmattered forms. The difference is that E. Thompson spells this enmattered precariousness out in terms of the *autopoiesis* of *constituent processes*, e.g. cells, whereas we expressed it, more generally, as the twofold insight that productive Autonomy must be inextricably enmattered because, first, precariousness is part of the definition of productive Autonomy and, second, we cannot understand precariousness without reference to matter (*as a principle of form*). However, insofar as both arguments rely on Jonas' reflections on *metabolism*, which E. Thompson basically identifies with the autopoietic organisation, this difference should not be overemphasised (we return to this shortly).

A more important difference is that *our* argument for the essential precariousness of productive Autonomy (itself a premise for arguing that productive Autonomy is inextricably enmattered) was not based on considerations restricted to enactivism. Rather, it proceeded from the constraints, particularly (C1), placed on *any* grounding of the notion of a productive system; a notion which, in turn, was based on a *general*, Anscombian conception of the notion of a productive process. It is only because, without precariousness, we cannot make sense of the notion of thick teleological explanation, of a process' explaining its *own* realisation, being its own source, that productive systems must be precarious (and therefore inextricably enmattered forms).

3.6.1 The Heart of the Matter: Life

What about *life*? Have we established that life, i.e. adaptive autopoiesis, is necessary for productive Autonomy, and hence necessary for agency? Almost. To cut the mustard we must show that autopoiesis, as it figures in (3)**, includes (or implies) adaptivity. Luckily, the odds are in our favour.

The flat reason is that, since adaptivity is an ingredient in the enactive definition of agency, it is a necessary feature of Autonomy_{AG} (i.e. productive Autonomy) *anyway*. A more informative reason is the following. As explained in Ch. 1, precariousness explains the important, non-accidental relation between autopoiesis and adaptivity. Since, to exist, a precarious system must constantly navigate the flow of matter and energy across its (organisational) boundary, it is *forced* to adaptively regulate its interactions with the environment. Adaptivity is not a contingent feature of precarious systems, but a necessary implication of their enmattered mode of being. If autopoiesis is to account for the precariousness of productive Autonomy, through its biochemical materiality, it had better retain this conceptual connection to adaptivity.

This point is closely related to Di Paolo et. al.'s emphasis on what they call the 'primordial tension' (2017: 134) internal to the notion of self-individuation (understood here in terms of a precarious, organisationally closed system). As they explain, '[t]he self-individuated system must tend to be self-enclosed to assert its distinctiveness as an individual, but it must also tend to be open to sustain its self-production as a far-from equilibrium system' (ibid). The resolution of this tension is its 'dialectical overcoming' (ibid.), achieved through 'adaptive regulation of the coupling with the environment' (ibid.). It's revealing that these considerations are presented as implications of the theory of *autopoiesis* and the notion of *biological* Autonomy (ibid.). Relatedly, Jonas is, unsurprisingly, cited as the source of inspiration – the notion of 'primordial tension' is an enactive spin-off of life's 'dialectical relation of *needful freedom* to matter' (Jonas 2001: 80, Original emphasis). These remarks also resonate with a central tenet of Moreno and Mossio's book *Biological Autonomy*, what they call 'the thermodynamic grounding of autonomy' (2015: Sect. 1.2). Roughly, the point is that the self-organising capacities of living systems cannot be understood without explicit reference to their material and energetic conditions as far-from-equilibrium systems (Moreno and Ruiz-Mirazo 1999). Although a full discussion of Moreno and Mossio's enquiry lies beyond our present topic, they clearly take this explicit reference to *matter conceptually* for our understanding of living systems. As they write:

> [A]n adequate understanding of biological organisation should reconcile form and matter, insofar as many fundamental features of biological organisation make sense, in [our] view, only in relation to the conditions of their realisation in nature. (2015: 7).

In sum, there are close conceptual ties between the enactive understanding of agency and the essentially material and precarious nature of *living* systems (cf. Moreno and Mossio 2015: Ch. 4). These aren't mere exegetical peculiarities. They point to the shared assumption that *there is something about life* – be it metabolism, the precariousness of biochemical macromolecules, Varela's 'processes of production', the autopoietic organisation, a primordial tension, or thermodynamic

instability – that is essential, *conceptually speaking*, for understanding the Autonomy of agents; something which, in all cases, is internally related to life's precarious relationship to matter.

Returning to our immediate topic, then, let us, following E. Thompson (2007: 151; 158), understand autopoiesis as including adaptivity. Then, since life is adaptive autopoiesis, productive Autonomy ($\text{Autonomy}_{\text{AG}}$) conceptually depends on life. Given our assumption that any agent is a productive system, this means that, conceptually speaking, life is necessary for agency. We can inscribe this conceptual connection in two slightly different ways into our definition.

> **(AA)**
> A system S is productive *iff* S's form is an Autonomous (organisationally closed) system of generic autopoietic processes.

> **(SA)**
> A system S is productive *iff* S's form is an Autonomous (organisationally closed) system of generic precarious processes, some of which are autopoietic *as a principle of form*.

According to **AA**, any productive system is *itself* an autopoietic system. All the processes constituting its closed, Autonomous organisation are precarious in an autopoietic way, i.e. because of their biochemical materiality. Hence **AA** for 'all-autopoietic'. According to **SA** ('some-autopoietic') a productive system *may* itself be wholly autopoietic, but it need not. The precariousness of its constituent processes may be partly grounded in the materiality of other kinds of processes which are not themselves autopoietic (this will become clearer below). However, we must emphasise that, on both **AA** and **SA**, the contribution of autopoietic materiality to our understanding of productive Autonomy *matters conceptually*. On both accounts, autopoiesis is necessary for productive Autonomy, not only in the sense that a productive system constitutively depends on autopoiesis (as in (3) and (3)*), but in the sense that autopoiesis figures *as a principle of form* in the *definition* of productive Autonomy. On both accounts, productive systems are inextricably autopoietic- and hence inextricably precarious, enmattered forms of Autonomy. Compare this to our earlier enactive definition of a productive system:

> A system S is productive *iff* S's form is an Autonomous (organisationally closed) system of generic precarious processes.

If either **SA** or **AA** coincide with this definition, then life will be necessary for agency (from an enactive perspective). Do we have grounds to believe that *neither* coincide?

3.6.2 Precariousness without Autopoiesis?

I am raising this question because, although our analysis *strongly suggests* that productive Autonomy conceptually depends on autopoiesis (life), there is logical space for a third possibility. For perhaps we can make sense of the precariousness

of certain productive systems in an entirely non-autopoietic way. After all, our enactive-style definition demands only that the productive processes constituting S's generic form be *precarious*. Following Jonas, we argued that precariousness conceptually depends on materiality, i.e. precarious forms are inextricably enmattered. But this leaves open the *kind* of materiality which substantiates this precariousness conceptually. Only if we can show that autopoietic, i.e. biochemical materiality is the *only* kind capable of accounting for the precariousness of productive systems can we conclude that life (autopoiesis) is necessary for productive Autonomy, and hence for agency. But might we not envisage a productively Autonomous system whose precarious materiality is completely divorced from the biochemical processes of life as we know it?

As anticipated in Ch. 1, the conceptual space in which this question finds a place seems to be the crux behind the enactive reluctance to claim that life is necessary for agency (Barandiaran et. al. 2009: 376). This reluctance is fuelled by recent developments illustrating that the concept of Autonomy can illuminate the organisational dynamics of systems which are strictly non-autopoietic. Indeed, Di Paolo et. al.'s (2017) main contribution, following Barandiaran (2008) and Barandiaran et. al. (2009), is a theory of *sensorimotor* agency, that is (roughly), a kind of agency which is realised in the behavioural domain, constituted by the circular, mutually reinforcing dependencies between sensation (perception) and motor-capacities (cf. Noë 2004). The chief biological basis for this kind of agency are the nervous and musculoskeletal system, hence why sensorimotor agency is (roughly) associated with the kingdom of *animals* (Di Paolo et. al. 2018: 44ff.) Similarly, Autonomous organisations are also thought to occur in the domain of social and linguistic interactions, as characterised (roughly) by the couplings between different agents (Cuffari et. al. 2015, Di Paolo et. al. 2018).

What essentially distinguishes these Autonomous organisations from autopoietic systems is their *domain* of organisation, determined by the kind of processes, e.g. nervous activity, sensorimotor engagements, interactive encounters and utterances, which realise organisational closure. However, this does not mean that these domain-specific systems are mutually exclusive. On the contrary, the crucial enactive thought is that 'a living body is the locus of different kinds of organisation' (Di Paolo et. al. 2018: 47). Organisms are, in Varela's words, 'a meshwork of selfless selves' (1991). Since a productive system, as defined, is a system which bears *a* generic Autonomous form, nothing precludes the possibility of one individual system embodying multiple, overlapping Autonomous forms (just as Beth can be an *animal*, a *human* and a *farmer*) (Di Paolo 2009).

Given this variety of Autonomous domains, might not a productive system be possible whose Autonomy is realised through organisationally closed processes whose precariousness is *detached* from autopoietic materiality, i.e. the biochemical domain of macromolecules and chemical reactions? Could we not 'bypass autopoiesis', as E. Thompson (2011: 215) wonders, and build a sensorimotor agent 'directly' (ibid.)? A satisfactory discussion of these questions would require another dissertation. Nevertheless, the foregoing analysis indicates, I think, a principled, conceptual restraint against stretching the notions of precariousness and productive Autonomy too far.

Our discussion of productive Autonomy as an inextricably precarious and hence enmattered form entails that the material features which account for

precariousness must be part, *as a principle of form*, of the definition of a productive system. Hence, to grasp the possibility of an entirely non-autopoietic productive system, we must be able to conceive of a system whose precariousness is grounded in a materiality which is not biochemical, which preconditions itself *without* 'something like a metabolism'. In the case of, say, a non-autopoietic, 'pure' sensorimotor agent, what would provide the relevant precarious materiality? How are sensorimotor processes, by themselves, *precarious*?

For autopoietic systems, there is a clear, thermodynamically grounded sense in which 'their constitutive structures and relations tend to decay and cannot exist except in the presence of the continuous regeneration of the whole organisation' (Moreno and Mossio 2015: 9). Deprive a bacterium of nutrients and it dies. Isolate a biochemical process from its intracellular medium and it peters out. Relatedly, each process within a living, autopoietic system occupies an *essential*, normatively significant role. It's always a sensible question to ask 'What's it for?' (Foot 2001: 31). The precariousness of such metabolic processes has a tangible grounding in their *material relevance* for upholding the system's organisation. There is thus a clear sense in which autopoietic systems are inextricably *enmattered* and *concretely* self-enabling. By comparison, how does it *matter*, quite literally, whether I grasp the mug handle with my pinky or my whole hand? The first of these engagements is 'worse' from a sensorimotor point of view. 'There is a smoother way to do it!', you might say. But what's the *point* of drinking coffee smoothly if the pinky does it? In what sense is smooth grasping a *precarious* process?

These critical questions are not meant to suggest that we cannot, to some extent, explain the precariousness and self-enablement of (sensorimotor) Autonomous organisations in *non*-autopoietic ways, e.g. in terms of their interdependent *coherence* (Barandiaran 2008, 2017). But they indicate the conceptual difficulties of grounding the precariousness of such networks materially, without, at some point, referring 'their normative or individuality conditions back to living organisation' (Barandiaran et. al. 2009: 376), i.e. without recognising their dependence on autopoietic processes. For if productive, precarious systems must be inextricably enmattered to explain self-enablement, materiality must figure somewhere, *as a principle of form*, in the definition of a productive system. Absent a clear understanding in which sensorimotor and other non-autopoietic processes, e.g. social interactions, are precarious, in a concrete, material sense, it seems that, for these non-autopoietic organisations to be properly enmattered, autopoietic materiality must figure somewhere, *as a principle of form*, in their definition. On this line of thought, any productive system will be dependent on having autopoietic parts which figure *as a principle of form* in our understanding of its Autonomy. Otherwise, we do not properly grasp how such a system is enmattered, and hence *productive* in a concrete, self-enabling sense. Our reflections thus support the view that our enactive definition of a productive system coincides (at least) with (**SA**), suggesting that productive Autonomy, and hence agency, conceptually depends on life.

3.6.3 The Emerging Picture

In sum, on the emerging view, [1] productive Autonomy is an inextricably precarious form of Autonomy. Since [2] precarious forms are inextricably enmattered and since [3] *prima facie* only autopoietic (i.e. metabolic, biochemical) materiality provides a clear, materially and thermodynamically based grounding for precariousness [4] autopoiesis must figure *as a principle of form* in the definition of productive Autonomy. This does not only mean that, to *construct* a productive system, one would ultimately need to rely on biochemical components (E. Thompson 2011: 215f.). For this position, as (3)* above, is compatible with holding that the Autonomous *form* of a productive system, what *defines* its productive Autonomy, can be understood independently of material aspects. On this view, autopoietic materiality is a *contingent* feature of productivity: it doesn't *matter conceptually*. Instead, [4] means that, *conceptually speaking*, we cannot grasp the Autonomous organisational form of a productive system apart from its relationship to autopoietic, biochemical material *as a principle of form*. This is the crux of saying that productive Autonomy is *inextricably enmattered* in an autopoietic way.

As with snubness, one consequence of this view is that there is no one common well-defined property, Autonomy*, which non-productive and productive systems share. Or, more precisely, if there is such a property, e.g. organisational closure, it is *not* productive Autonomy (just as the geometrical concavity which bandy legs and snub noses share is, on Aristotle's view, not *snubness*). This is a virtue of the view because it keeps the Autonomy of productive systems conceptually tied to precarious, autopoietic materiality *as a principle of form* and thereby preserves the features which *explain* how such a system is productive in the first place. For precariousness explains how such a system is self-enabling, and thus accounts for the notion of *thick* teleological explanation underwriting the notion of a productive process or system. If we drove precarious, autopoietic materiality out of our *definition* of a productive system, the concept would lose this explanatory power. Since *productive* Autonomy coincides with the Autonomy of agents, i.e. $\text{Autonomy}_{\text{AG}}$, this means that, where $\text{Autonomy}_{\text{AG}}$ is concerned, it is not enough to *add* that 'closure obtains under precarious conditions' (Di Paolo et. al. 2018: 25). Precariousness must be *part* of the definition of the Autonomy of agents, as indeed it is in Di Paolo et. al.'s (2017: 127) definition of agency.

The emerging view also preserves 'the Anscombian thought'. That is, the metabolic precariousness of the component processes of a productive system, which explains the sense in which these processes are *productive*, is only intelligible against the background of the Autonomous, closed *form* which unites its generic constituent processes into a whole. Thus, to call a particular process *productive* presupposes the 'wider context' of the Autonomous network of generic precarious processes to which it belongs, contributes and on which it depends. Crucially, however, given our previous point, our understanding of the Autonomous *form* of this network is itself conceptually dependent on its autopoietic materiality. In that sense, productive Autonomy is not only an inextricably enmattered form, its precarious, autopoietic materiality is also an 'inextricably enformed' kind of materiality. In a productive system, form and matter constitute a unity which cannot be severed without losing sight of what makes the system

productive. Before concluding, let me speculate about one final idea which, I think, is missing from the picture.

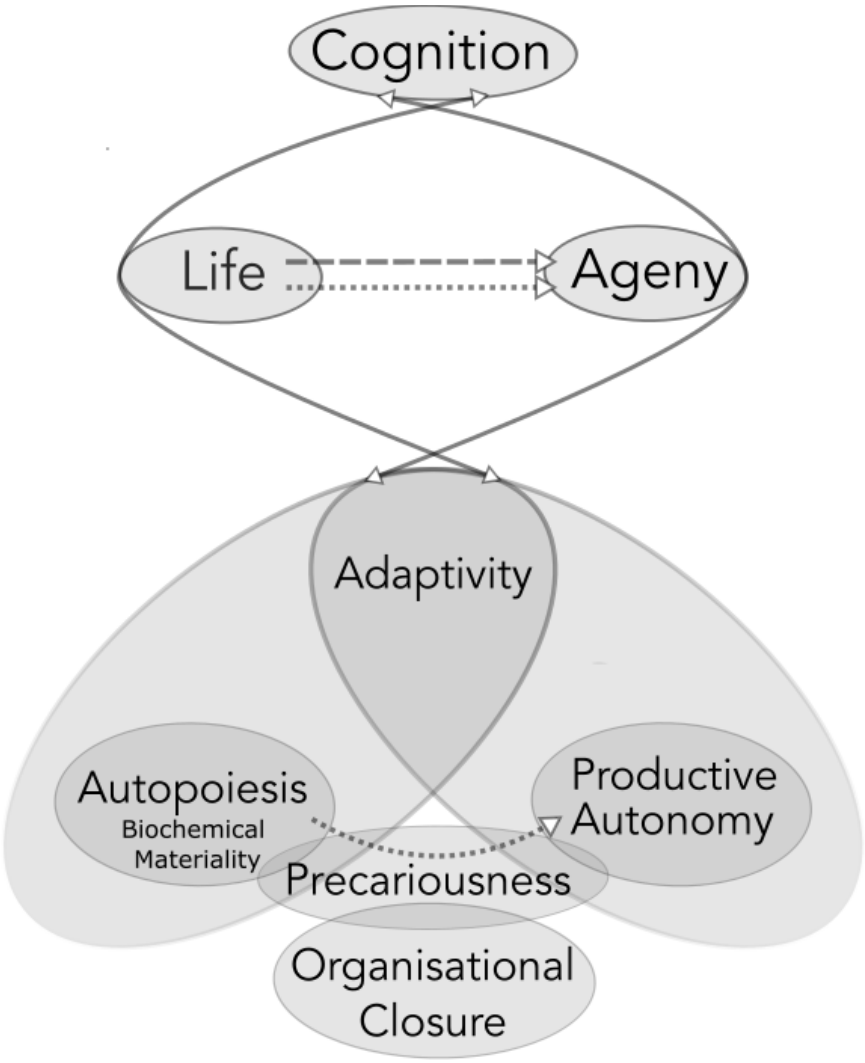


Figure 2 The continuity of life, agency and cognition. Conceptual relations are represented as follows: dotted lines indicate *necessity*, dashed lines *sufficiency* and bidirectional, bold lines *necessity and sufficiency* (equivalence). On the present view, life, agency and cognition are not (necessarily) equivalent but conceptually entangled (represented by the bold lines' arrows connecting their conceptual fringes into a closed circle). In particular, life is necessary for agency since (a) productive Autonomy conceptually depends on precariousness and (b) precariousness conceptually depends on the biochemical materiality of autopoiesis. (a) and (b) are represented by the dotted line between autopoiesis and productive Autonomy as well as the conceptual overlap of both concepts with precariousness, which in turn overlaps with organisational closure.

3.7 The Missing Piece: Internal Identity

Autonomy, we argued, cannot account for individuality in the sense of numerical identity if it is to provide a grounding for normativity. To provide such a grounding, we said, (productive) Autonomy must be understood generically as defining a *kind* of organisation, corresponding to a *class* of systems. Analogously to life-forms, all individual members of this class are governed by the same normative standard, determined by their shared Autonomous organisation. But in what sense exactly are these members *individuals*?

A similar issue emerges from the observation, anticipated earlier, that our proposal counts processes as productive which, intuitively, are not *actions*. The photosynthesis of a mitochondrion within a particular cyanobacterium, for instance, is an instance of a generic process within the precarious, autopoietic organisation of processes which constitutes the generic Autonomous form which the mitochondrion exemplifies. Accordingly, *the mitochondrion is photosynthesising* is the description of a productive process. Relatedly, *the mitochondrion is photosynthesising because it builds carbohydrates* expresses a *thick* teleological explanation. For the building of carbohydrates is itself a precarious process which *preconditions itself* by contributing to the bacterium's metabolism on which the mitochondrion depends. But is it right to describe photosynthesis as *coming from the mitochondrium*? Perhaps. Nonetheless, it seems wrong to call the mitochondrion's activity an *action*. The mitochondrion, insofar as it is *part* of the bacterium, does not seem to be the kind of thing which can perform actions.

Something is missing to distinguish *actions* from productive processes generally, something registered by the following point: actions are done by *individuals*, not by their *parts* or by their *kind*. This point is contained in Di Paolo et. al.'s self-individuation condition for agency, which Autonomy was to subserve. But if we think of Autonomy generically, the resulting notion of individuality mirrors M. Thompson's notion of an individual organism as *anything that bears a life-form*. However, as noted earlier, this notion wrongly implies that *the heart of my neighbour's cat* is an individual organism. Indeed, it seems to treat all instances of an Autonomous-/life-form as *the same individual.* One might counter: but these instances can be *distinguished* from one another, e.g. through their different spatiotemporal position. Indeed, the each satisfy Leibniz's law, i.e. they are numerical individuals – you can count them after all! But this, I think, misunderstands the *kind* of identity required for agency. When individual agents do things, they do so not just as members of their kind, but as *themselves,* where this means *more* than just *being distinct.* Jonas expresses this point clearly in relation to life.

> […] to understand the individuality of the organism, we have to transcend even the aspect of form […] we have now to take seriously the reflexive pronoun »itself«, i.e., to place the principle of identity in that which exercises this freedom of form – something which abstract form as such cannot do […] the ontological individual requires, behind the continuity of form which it shares with the wave or the flame, *internal* identity as the subject of its existing *in actu*. (Jonas 2016: 54f. Original emphasis.)

> […] form as such shows wholeness [identity] rather than provides it, and its persistence under the conditions of metabolic turnover qualifies for the attribution of individuality only if behind it is assumed a principle of identity operative in the mere morphological fact. (Jonas 2016: 60).

Jonas is often (and understandably) read as identifying an organism's identity with the *form* maintained throughout the constant exchange of matter. But Jonas rejects such a formal account.[22] As he says, the true, ontological individuality of the organism transcends 'even the aspect of form'. It is an *internal* identity, a *self*, a 'wholeness [which] is self-integrating in active performance' (Jonas 2001: 79).

Jonas, and his enactive followers, have been criticized for employing terms like 'selfhood' or 'self-concern' to even (relatively) simple organisms like bacteria (De Jesus 2016, Ward and Villalobos 2016a, Kee 2021). However, without entering this debate, it seems that the notion of *self*, if understood in the minimal sense of an *actively self-integrating identity* or *wholeness*, is appropriate from the perspective of explaining the *agency* of productive systems. It encapsulates the insight that, when we call something and *agent*, we treat it as an individual which acts for and by itself and thereby manifests a distinction between self-movement and other-movement ('movement' in the broad sense of a *process* as explained in Ch. 2) (cf. Hornsby 2004). This sense of individuality – let us, following Jonas, call it *internal identity* – seems not to be captured by the notion of form alone, nor, indeed, is it captured by Leibniz's Law or numerical identity. To be a self-integrating individual capable of performing actions is not only a matter of maintaining a certain form of organisation or of being distinct. It is to be a *unity* capable of exerting an ontological claim as a causally efficacious, purposive being within the fabric of reality.

This notion of a self-integrating wholeness or internal identity is vague. Nevertheless, it indicates the *kind* of identity which, I think, underlies the distinction between an *individual* as opposed to *its* physical environment, *its* parts and *its* kind, and it is the kind of identity required to distinguish actions from productive processes generally. Recognising this internal identity in even simple organisms is not to attribute to them a self-conscious kind of selfhood. It is to acknowledge their wholeness as individuals with an identity which belongs to them, and no one else, and whose possibility and reality depends on their own doing. Of course, as Jonas himself was aware (2001: 82; 2016: 54ff.; cf. Prokop, forthcoming), this identity is more a conjecture than an explanation. Indeed, it was perhaps what Maturana and Varela originally had in mind when emphasising the 'self-asserting capacity of living systems to maintain their identity' (1973: 73), i.e. life's *autonomy*, understood in the intuitive sense with which we began, and which autopoiesis and Autonomy were intended to capture. If our previous considerations are right, however, and Autonomy is to be understood generically, it is doubtful whether these concepts adequately capture the notion of internal identity. This is not, I think, necessarily a *problem* for the enactive approach. Indeed, one might be sceptical whether so much as a *theory* of internal identity is possible or required. Perhaps recognising it is simply to grant the organism a 'minimal freedom', in Bohr's words, 'just large enough to permit it, so to say, to hide its ultimate secrets from us.' (1999: 34). At any rate, our previous discussion, if anything, emphasises the need to understand this identity in its relation to productive existence, and thus

[22] See Barbaras (2010), Coyne (2021: 48ff.) and Prokop (forthcoming) for discussion.

as residing in the essentially teleological, projective and progressively unfolding sphere of being we have been investigating. From this perspective, to treat the identity of living things as an *object* of inquiry on a par with the identity of non-living things is already to fundamentally misunderstand its nature.

Conclusion

Productive being, the precarious existence of productive systems, is indeed 'a remarkable mode of being' (Jonas 2001: 75). In relying on the exchange of matter with its environment, a productive system is constantly engaged in actively maintaining its integrity. In this dependence on materiality, a productive system, I argued, embodies an inextricably enmattered form of Autonomy, which cannot be understood without explicit reference to matter *as a principle of form*.

The inextricably enmattered, precarious form of productive beings grounds their capacity to engage in productive processes. Such processes, I argued, are not merely the objects of intentional descriptions, although the teleological-explanatory form of the intentional order is essential to understanding their progressive, end-directed structure. They are processes which come from their subjects and explain their own realisation, in a non-accidental manner. This applicability of thick teleological explanations to productive processes lies in their contribution to, and dependence on, the maintenance of the precarious Autonomous form of their subject – a form which, because it embodies this organisationally closed interdependence of productive processes, is self-enabling. The capacity of productive subjects for productive processes is, however, also a necessity. Due to their precarious, enmattered nature, they cannot but perpetually reproduce themselves. This need for active maintenance connects to the internal normativity which governs the processes of productive systems. To explain this normativity, I argued, their Autonomy must be conceived generically, as defining a general kind of organisation of processes, similar to the notion of a life-form.

From an enactive perspective, I suggested, it is hard to properly explain the precariousness of productive systems without recognising their dependence on the biochemical, precarious materiality of autopoiesis. Autopoiesis (including adaptivity) thus *matters conceptually* for our understanding of productive systems and their teleological mode of being. Since agents are productive systems – we plausibly assumed – this means that life (adaptive autopoiesis) is necessary for agency, conceptually speaking. Furthermore, since, as explained in Ch. 1, sense-making (basic cognition) is dependent on agency, this insight also answers question (B): life is necessary for cognition (although this claim certainly requires more attention than it received here). In sum, on the present view, the deep continuity of life and mind becomes entangled in the continuity between life and agency, grounded in the dependence of action on productive Autonomy on life.

The claim that the precariousness of productive systems cannot be understood without explicit reference to autopoietic, biochemical materiality occasions, I suspect, the most doubt in my argument. Perhaps it is possible to understand precariousness in non-autopoietic terms, to conceive of an agent whose self-enabling capacities do not derive from the biochemical materiality of life as we know it. Investigating these possibilities is a task for future research. Nevertheless,

on the Jonasian-Aristotelian line of thought developed here, precariousness remains an inextricably enmattered property. Hence, to make sense of the essential precariousness of agents *qua* productive systems, materiality must figure, *as a principle of form*, in the definition of their Autonomy. If, plausibly, we understand *autonomy*, i.e. life's self-integrating capacity, with which we began, as an essentially productive capacity, then Maturana and Varela were either wrong to assume that the autonomy of living systems can be understood in purely organisational terms 'whichever the nature of their components' (1973: 76) or, more likely, they themselves implicitly understood autopoiesis, the organisation of the living, as an inextricably enmattered form of being. Life and action matter, after all.

Bibliography

Anscombe, G. E. M. (2000). *Intention*. 2nd Ed. Cambridge, MA: Harvard University Press.

Aquinas, T. (1912). *The 'Summa Theologica'*, Part I, Fathers of the English Dominican Province (trans.), 3rd. Number, London: R&T Washbourne Ltd.

Aristotle (2016). *De Anima*. C. Shields (trans.). Oxford: Oxford University Press.

Barandiaran, X. E. (2008). *Mental life. A naturalized approach to the autonomy of cognitive agents*. Unpublished doctoral dissertation, University of the Basque Country,Spain.https://xabierbarandiaran.files.wordpress.com/2010/06/barandiaran_-_2008_-_phd_thesis.pdf (last access 22.08.2022).

Barandiaran, X. E. (2017). Autonomy and Enactivism: Towards a Theory of Sensorimotor Autonomous Agency. *Topoi*, Vol. 36, pp. 409-430.

Barandiaran, X. E., Di Paolo, E., Rohde, M. (2009). Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-temporality in Action. *Adaptive Behaviour*, Vol. 17, pp. 367-386.

Barandiaran, X. E. and Di Paolo, E. (2014). A genealogical map of the concept of habit. *Frontiers in Human Neuroscience*, Vol. 8, pp. 1-7.

Barbaras, R. (2010). Life and exteriority: The problem of metabolism. In J. Stewart, O. Gapenne, & E. A. Di Paolo (eds.). *Enaction: Toward a new paradigm for cognitive science*. Cambridge, MA: MIT Press, pp. 89–122.

Barrett, N. F. (2017). The Normative Turn in Enactive Theory: An Examination of Its Roots and Implications. *Topoi*, Vol. 36, pp. 431-443.

Bohr, N. (1999). *Collected Works*. F. Aaserud and D. Favrholdt (eds.), Vol. 10, Amsterdam: Elsevier.

Boyle, M. and Lavin, D. (2010). Goodness and Desire. In S. Tenenbaum (ed.). *Desire, Practical Reason, and the Good*. Oxford: Oxford University Press

Capra, F. and Luisi, P. L. (2014). *The Systems View of Life: A Unifying Vision*. Cambridge: Cambridge University Press.

Charles, D. (2021). *The Undivided Self: Aristotle and the 'Mind-Body' Problem*. Oxford: Oxford University Press.

Clark, A. (2001). *Mindware: an introduction to the philosophy of cognitive science.* New York: Oxford University Press.

Coope, U. (2007). Aristotle on Action. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 2007, Vol. 81, pp. 109-138.

Coyne, L. (2021). *Hans Jonas: Life, Technology and the Horizons of Responsibility*. London: Bloomsbury Academic.

Cuffari, E. C., Di Paolo, E., De Jaegher, H. (2015). From participatory sense-making to language: there and back again. *Phenomenology and the Cognitive Sciences*, Vol. 14, pp. 1089-1125.

Davidson, D. (1982). Rational Animals. *Dialectica*, Vol. 36, pp. 317-327.

Davidson, D. (2001). *Essays on Actions and Events*. 2nd Ed. Oxford: Clarendon Press.

De Jaegher, H. and Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, Vol. 6, pp. 485-507.

De Jesus, P. (2016). Autopoietic enactivism, phenomenology and the deep continuity between life and mind. *Phenomenology and the Cognitive Sciences,* Vol. 15, pp. 265–289.

Dennett, D. (1991). Real Patterns. *The Journal of Philosophy*, Vol. 88, pp. 27-51.

Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, Vol. 4, pp. 429-252.

Di Paolo, E. (2009). Extended Life. *Topoi*, Vol. 28, pp. 9-21.

Di Paolo, E. (2018). The Enactive Conception of Life. In A. Newen, L. De Bruin, and S. Gallagher (eds.). *The Oxford Handbook of 4E Cognition*, Oxford: Oxford University Press, pp. 71-94

Di Paolo, E., Rohde, M. and De Jaegher, H. (2010). Horizons for the Enactive Mind: Values, Social Interaction, and Play. In J. Stewart, O. Gapenne, and E. Di Paolo (eds.). *Enaction: Towards a New Paradigm of Cognitive Science,* Cambridge, MA: MIT Press, pp. 33–87.

Di Paolo, E. and Thompson, E. (2014). The Enactive Approach. In L. Shapiro (ed.). *The Routledge Handbook of Embodied Cognition.* Abingdon: Routledge, pp. 68-78.

Di Paolo, E., Buhrmann, T. and Barandiaran, X. E. (2017). *Sensorimotor Life: An Enactive Proposal*. Oxford: Oxford University Press.

Di Paolo, E., Cuffari, E. C. and De Jaegher, H. (2018). *Linguistic Bodies*. Cambridge, MA: MIT Press.

Foot, P. (2001). *Natural Goodness.* New York: Oxford University Press.

Ford, A. (2011). Action and Generality. In A. Ford, J. Hornsby and F. Stoutland (eds.). *Essays on Anscombe's Intention*. Cambridge, MA: Harvard University Press, pp. 76-104.

Ford, A., Hornsby, J. and Stoutland, F. (eds.) (2011). *Essays on Anscombe's Intention.* Cambridge, MA: Harvard University Press.

Frege, G. (1891/1997). Function and Concept. Reprinted in M. Beaney (Ed.), *The Frege reader*. Oxford: Blackwell, pp. 130-148

Frege, G. (1891–1895/1997). Comments on Sinn and Bedeutung. Reprinted in M. Beaney (Ed.), *The Frege reader*. Oxford: Blackwell, pp. 172–180

Frege, G. (1892a/1997). On Concept and Object. Reprinted in M. Beaney (Ed.), *The Frege reader.* Oxford: Blackwell, pp. 181-193.

Frege, G. (1892b/1997). On Sinn and Bedeutung. Reprinted in M. Beaney (ed.), *The Frege reader*. Oxford: Blackwell, pp. 151-171.

Frege, G. (1918/1997). Thought. Reprinted in M. Beaney (ed.), *The Frege reader*. Oxford: Blackwell, pp. 325-345.

Froese, T. (2017). Life is Precarious Because it is Precarious: Individuality, Mortality, and the Problem of Meaning. In G. Dodig-Crnkovic & R. Giovagnoli (eds.). *Representation and Reality in Humans, Other Living Organisms and Intelligent Machines.* Heidelberg: Springer, pp. 33-50.

Froese, T. and Ziemke, T. (2009). Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence*, Vol. 173, pp. 466-500.

Fromm, E. (1962). *The Art of Loving*. London: Unwin Books.

Godfrey-Smith, P. (1994). Spencer and Dewey on life and mind. In Brooks, R. and Maes, P. (eds.). *Artifical Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*. Cambridge, MA.: MIT Press.

Godfrey-Smith, P. (2016). Mind, Matter, and Metabolism. *The Journal of Philosophy*, Vol. 113, pp. 481-506.

Haase, M. (2018). Practically Self-Conscious Life. In J. Hacker-Wright (ed.), *Philippa Foot on Goodness and Virtue.* Cham, Switzerland: Palgrave Macmillan, pp. 85-126.

Hall, S. (2022). Webb Telescope Detects Carbon Dioxide on Exoplanet. *Nature,* Vol. 609, pp. 229-230.

Haugeland, J. (1998) *Having Thought: Essays in the metaphysics of mind.* Cambridge, MA: Harvard University Press.

Hornsby, J. (2004). Agency and Actions. *Royal Institute of Philosophy Supplement,* Vol. *55*, pp. 1-23.

Hornsby, J. (2011). Actions in Their Circumstances. In A. Ford, J. Hornsby and F. Stoutland (eds.). *Essays on Anscombe's Intention*. Cambridge, MA: Harvard University Press, pp. 105-127.

Hverven, S. and Netland, T. (2021). Projection or encounter? Investigating Hans Jonas' case for natural teleology. *Phenomenology and the Cognitive Sciences*, published online, doi: https://doi.org/10.1007/s11097-021-09748-9

Jonas, H. (1968). Biological Foundations of Individuality. *International Philosophical Quarterly*, Vol. 8, pp. 231-251.

Jonas, H. (1996). *Mortality and Morality: A Search for the Good After Auschwitz*, L. Vogel (ed.). Evanston, IL: Northwestern University Press.

Jonas, H. (2001). *The Phenomenon of Life: Toward a Philosophical Biology*. Evanston, IL: Northwestern University Press.

Jonas, H. (2016). The Basic Mode of Organic Existence: Metabolism. In J. O. Beckers und F. Preußger (eds.), *Organism and Freedom, Online-Appendix zu Band I, 1 der Kritischen Gesamtausgabe der Werke von Hans Jonas*, Ch. II.

JWST Transiting Exoplanet Community Early Release Science Team. (2022). Identification of carbon dioxide in an exoplanet atmosphere. *Nature*, https://doi.org/10.1038/s41586-022-05269-w

Kant, I. (2000). *Critique of the Power of Judgement*. P. Guyer, E. Matthews (trans.), Cambridge: Cambridge University Press.

Kee, H. (2021). Phenomenology and naturalism in autopoietic and radical enactivism: exploring sense-making and continuity from the top down. *Synthese*, Vol. 198, Suppl. 9, pp. S2323-S2343.

Leslie and Lerner (2022). Generic Generalizations. *The Stanford Encyclopedia of Philosophy.* Fall 2022 Edition, Edward N. Zalta (ed.), forthcoming URL = https://plato.stanford.edu/archives/fall2022/entries/generics/

Lewens, T. (2012). Human Nature: The Very Idea. *Philosophy & Technology*, Vol. 25, pp. 459-474.

Lewens, T. (2020). Species Natures: A Critique of Neo-Aristotelian Ethics. *The Philosophical Quarterly*, Vol. 70, pp. 480-501.

Lewis, C. S. (1967/2013). *Studies in Words*. 2nd Ed. Canto Classics. Cambridge: Cambridge University Press.

Maturana, H. R. (1970). Biology of Cognition. In H. R. Maturana and F. J. Varela (1980). *Autopoiesis and Cognition: The Realization of the Living*, Dordrecht, Holland: D. Reidel Publishing Company

Maturana, H. (1975). The organization of the living: A theory of the living organization. *Intl. J. Man-Machine Studies*, Vol. 7, pp. 313–332.

Maturana, H. R. and Varela, F. J. (1973). Autopoiesis: The Organization of the Living. In H. R. Maturana and F. J. Varela (1980). *Autopoiesis and Cognition: The Realization of the Living*, Dordrecht, Holland: D. Reidel Publishing Company

Maturana, H. R. and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*, Dordrecht, Holland: D. Reidel Publishing Company

Moreno, A. and Ruiz-Mirazo, K. (1999). Metabolism and the problem of its universalization. *BioSystems*, Vol. 49, pp. 45-61.

Moreno, A. and Mossio, M. (2015). *Biological Autonomy: A Philosophical and Theoretical Enquiry.* Dordrecht: Springer.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

Noonan, H. and Curtis, B. (2022). Identity. *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), E. N. Zalta and U. Nodelman (eds.), forthcoming, https://plato.stanford.edu/archives/fall2022/entries/identity/

Prokop, M. (forthcoming). Hans Jonas and the Phenomenological Continuity of Life and Mind. *Phenomenology and the Cognitive Sciences.*

Ramsey, F. P. (1925). Universals. Reprinted in D. H. Mellor (ed.), *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, London and Henley: Routledge & Kegan Paul, pp. 17-39.

Ross, D. (1964). *Aristotle*. 5th Ed. University Paperbacks. London: Methuen & Co Ltd.

Rödl, S. (2007). *Self-Consciousness*. Cambridge, MA: Harvard University Press.

Rödl, S. (2012). *Categories of the Temporal: An Inquiry into the Forms of the Finite Intellect.* S. Salewski (trans.), Cambridge, MA: Harvard University Press.

Rudrauf, D., Lutz, A., Cosmelli, D. Lachaux, J.-P., and Le Van Quyen, M. (2003). From autopoiesis to neurophenomenology: Francisco Varela's exploration of the biophysics of being. *Biological Research*, Vol. 36, pp. 27-66.

Rushforth, K. (1980). *The Mitchell Beazley pocket guide to Trees*. London: Mitchell Beazley Publishers Ltd.

Schlosser, M. (2019). Agency. *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2019/entries/agency/

Schulz, H. N., Brinkhoff, T., Ferdelman, T. G., Hernández Mariné, M., Teske, A. and Jørgensen, B. B. (1999). Dense Populations of Giant Sulfur Bacterium in Namibian Shelf Sediments. *Science*, Vol. 284, pp. 493-495.

Sosa, E. and Tooley, M. (eds.) (1993). *Causation*. Oxford: Oxford University Press.

Stoutland, F. (2011). Introduction: Anscombe's *Intention* in Context. In A. Ford, J. Hornsby and F. Stoutland (eds.). *Essays on Anscombe's Intention*. Cambridge, MA: Harvard University Press, pp. 1-22.

Thompson, E. (2007). *Mind in Life*. Cambridge, MA: Harvard University Press.

Thompson, E. (2011). Reply to Commentaries. *Journal of Consciousness Studies*, Vol. 18, pp. 176-223.

Thompson, E. and Stapleton, M. (2009). Making Sense of Sense-Making. *Topoi*, Vol. 28, pp. 23-30.

Thompson, M. (1995). The Representation of Life. In R. Hursthouse, G. Lawrence and W. Quinn (eds.). *Virtues and Reasons: Philippa Foot and Moral Theory. Essays in Honour of Philippa Foot.* Oxford: Clarendon Press, pp. 247-269.

Thompson, M. (2004). Apprehending Human Form. *Royal Institute of Philosophy Supplements*, Vol. 54, pp. 47-74.

Thompson, M. (2008). *Life and Action: Elementary Structures of Practice and Practical Thought.* Cambridge, MA: Harvard University Press.

Trueman, R. (2015). The concept *horse* with no name. *Philosophical Studies*, Vol. 172, pp. 1889-1906.

Varela, F. J. (1979). *Principles of Biological Autonomy.* New York: Elsevier North Holland.

Varela, F. J. (1991). Organism: A Meshwork of Selfless Selves. In A.I. Tauber (ed.) *Organism and the Origins of Self,* Boston Studies in the Philosophy of Science, Vol. 129, Dordrecht: Kluwer Academic Publishers, pp. 79-107.

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

Varela, F. J., Maturana, H., Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterisation and a model. *Biosystems*, Vol. 5, pp. 187-196.

Varela, F. J., Thompson, E., Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience,* Cambridge, MA: MIT Press.

Ward, D. and Villalobos, M. (2015). Living Systems: Autonomy, Autopoiesis and Enaction. *Philosophy & Technology,* Vol. 28, pp. 225-239

Ward, D. and Villalobos, M. (2016a). Lived Experience and Cognitive Science: Reappraising Enactivism's Jonasian Turn. *Constructivist Foundations*, Vol. 11, pp. 204-212.

Ward, D. and Villalobos, M. (2016b). Authors' Response: Enactivism, Cognitive Science, and the Jonasian Inference. *Constructivist Foundations*, Vol. 11, pp. 228-233.

Ward, D., Silverman, D. and Villalobos, M. (2017). Introduction: The Varieties of Enactivism. *Topoi*, Vol. 36, pp. 365-375.

Weber, A. and Varela, F. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences,* Vol. 1, pp. 97-125.

Wheeler (1997). Cognition's coming home: The reunion of life and mind. In Husbands, P. and Harvey, I. (eds.). *Proceedings of the 4th European Conference on Artificial Life*, Cambridge, MA: MIT Press, pp. 10-19.

Wheeler, M. (2011). Mind in Life or Life in Mind? Making Sense of Deep Continuity. *Journal of Consciousness Studies*, Vol. 18, pp. 148-168.

Wittgenstein, L. (1953). *Philosophical Investigations*. G. E. M. Anscombe (trans.). Oxford: Basil Blackwell.

Wright, C. (1998). Why Frege Does Not Deserve His Grain of Salt. In B. Hale and C. Wright (eds.). *The Reason's Proper Study*. Oxford: Oxford University Press, pp. 72-90.