Title page

This thesis is presented for the degree of Doctor of Philosophy

Thesis title:

Diachronic Metaphysical Building Relations: Towards the Metaphysics of Extended Cognition

Name and degrees of candidate:

Michael David Kirchhoff

BA in philosophy and general semiotics (Aarhus University, Denmark); MA in philosophy (Aarhus University, Denmark)

Affiliation:

Department of Philosophy, Macquarie University, Sydney, NSW, Australia

Submission date:

Word count:

87.266

Contents

Thesis abstract

In the thesis I offer an analysis of the metaphysical underpinnings of the extended cognition thesis via an examination of standard views of metaphysical building (or, dependence) relations.

In summary form, the extended cognition thesis is a view put forth in naturalistic philosophy of mind stating that the physical basis of cognitive processes and cognitive processing may, in the right circumstances, be distributed across neural, bodily, and environmental vehicles. As such, the extended cognition thesis breaks substantially with the still widely held view in cognitive science and philosophy of mind, namely that cognitive processes and cognitive processing take place within the skin-and-skull of individual organisms.

The standard view of metaphysical building relations can be expressed as the conjunction of two theses. First, that a metaphysical building relation – such as composition, constitution, realization, supervenience or emergence – is a relation of ontological dependence, because if a metaphysical building relation holds between X (or the Xs) and Y, then it is in virtue of X (or the Xs) that Y exists. Second, metaphysical building relations are synchronic (durationless) relations of ontological dependence.

In the thesis, I propose an alternative *diachronic* framework by which to extend the standard synchronic accounts of metaphysical dependence relations, and by which to reformulate the metaphysical foundation of the extended cognition thesis. The project fills an important gap between analytical metaphysics (in particular, the metaphysics of dependence relations) and naturalistic philosophy of mind (especially the extended cognition thesis). To my knowledge there has been no attempt to establish a robust diachronic account of metaphysical building (or,

dependence) relations such as, e.g., composition and constitution. However, this is precisely what I argue is required to properly advance and ground the metaphysics of extended cognition. Ultimately, my aim of reformulating the metaphysics of extended cognition consists in taking several steps toward a third-wave of extended cognition.

Statement by candidate

I certify that the work in this thesis entitled “Diachronic Metaphysical Building Relations: Towards the Metaphysics of Extended Cognition” has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and preparation of the thesis itself have been appropriately acknowledged.

In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature:

Date:

Michael David Kirchhoff (42604524)

Acknowledgements

I would like to thank the Australian Research Council (ARC) and in particular Richard Menary and David Simpson, because this PhD project has been funded by an ARC PhD scholarship, as part of an ARC Discovery Project Grant DP1095109 awarded to Richard Menary and David Simpson. I am also grateful to Macquarie University for providing me with an International Macquarie University Research Excellence Scholarship (No. 2011180), while I was transferring my degree from the University of Wollongong to Macquarie University, for tuition fee support during my two years of postgraduate studies at Macquarie University, and for enabling me to carry out my research in a stimulating research environment.

Thanks are due to many people for discussion of related issues and comments on earlier drafts, but especially to Richard Menary, John Sutton, Will Newsome, Julian Kiverstein, Shaun Gallagher, Dan Hutto, Ken Aizawa, Chris Winch, David Simpson, Dan Zahavi, Johanna Seibt, Henning Høgh Laursen, Patrick McGivern, Neil McDonnell, and to various members of the audiences on several occasions when some of the material in the thesis has been presented. Special thanks to Karen Bennett for the permission to read and cite her forthcoming book *Making Things Up* with Oxford University Press. I am also grateful to my former teacher and mentor, Steen Wackerhausen, without whom my interest in distributed and extended cognition might never have evolved. Finally, and much closer to home, thanks to Casey Gibson and the Gibson family for their support and for allowing me to spend many hours of efficient and undisturbed writing in their beautiful home in Currarong, NSW, Australia.

Parts of chapters 1 and 8 are based on my treatment of the constitution relation and extended cognition, to be found in my papers "Extended Cognition & the

Causal-Constitutive Fallacy: In Search for a Diachronic and Dynamical Conception of Constitution” published in *Philosophy and Phenomenological Research* (DOI: 10.1111/phpr.12039) and “Extended Cognition & Constitution: Re-evaluating the Constitutive Claim of Extended Cognition” published in *Philosophical Psychology* (DOI: 10.1080/09515089.2012.724394). Parts of chapter 4 are based on my treatment of the recent debate between Clark and Hutchins on the timescales over which cognition and culture may interact, to be found in my papers “Extended Cognition and Fixed Properties: Steps to a Third-Wave Version of Extended Cognition” published in *Phenomenology and the Cognitive Sciences* (2012, 11(2), 287-308), “Distributed Cognitive Agency in Virtue Epistemology” published in *Philosophical Explorations* (2012, 15(2), 165-180), and a paper that is currently under review in *Phenomenology and the Cognitive Sciences*. Parts of chapter 6 are currently under review in *Synthese*, while parts of chapter 7 are based on my paper “In Search for Ontological Emergence: Diachronic, but Non-Supervenient” published in *Axiomathes* (DOI: 10.1007/s10516-013-9214-7). Thanks to the referees from each of these journals for valuable comments and suggestions for improvement. Thanks also to the examiners of the thesis for suggestions for improvement.

For my mother and grandparents with thanks

1. Introduction: DIACHRONIC versus SYNCHRONIC

My general concern in the thesis is with the idea that cognitive processes and cognitive systems may be instantiated by elements distributed beyond the brain and/or whole body of organisms to include environmental components. As one of the key exponents of this view has expressed this ontological claim: "Cognitive processes are not located exclusively inside the skin of cognizing organisms." (Rowlands 1999, p. 22; see also Clark 1997, 1998a, 2003, 2008; Clark & Chalmers 1998; Menary 2007, 2010b; Sutton 2010; Wheeler 2010; Wilson 1994, 2004a). This is the *extended cognition thesis* (EC). It provides arguments for the view that the boundaries of cognition are not bounded by either the brain or the organismic body but may, in the right circumstances, be broadened to include environmental resources (e.g., artifacts, people, social institutions, practices, and so on).

The primary question that shall occupy me is the following: if the relationship between the physical machinery of cognition, on the one hand, and cognition, on the other, is one of non-identity (which is the received view in contemporary naturalistic philosophy of mind), what is the relationship between cognition[1] and its physical substrate? Is the relationship a supervenience relation? Is the relationship a realization relation? Constitution? Emergence? Or, might it be a composition relation? Independent of which specific relation one opts for, these relations are all

[1] Here and throughout, I am using "cognition" as a catch-all term covering (i) processes such as remembering, learning, problem solving, perception, and so on, (ii) states such as beliefs and desires, and (iii) consciousness. However, I will be focusing predominantly on processes such as remembering, problem solving, perception, and learning.

relations of metaphysical dependence. I refer to these different relations as METAPHYSICAL BUILDING RELATIONS[2].

There are (at least) two different modes in which these metaphysical building relations may be put to use. The first is *proprietary*, which specifies that a relation like realization, say, holds only between "the physical" and "the mental". The second is *generic*. In contrast to the proprietary sense of realization, the generic mode of realization implies that the relation of realization is a general relation that holds independently of the nature of the phenomena in question (Wilson 2004a, p. 100)[3].

In the philosophy of mind, arguably the most familiar view of the relationship between mind and body comes from a position pioneered by Putnam (1960/[1975]). Putnam formulated the relationship between the mental and the physical as one of realization. According to this view, mental states or properties are realized by physical states or properties of the brain but are neither identical with nor reducible to such brain states or properties. Thus, Putnam used realization in its proprietary mode in order to express that the relationship between mind and body is one of realization. Supervenience has also been used to express the idea of the mental and the physical as being in a relation of dependence, from the physical to the mental, yet without that dependence relation being an identity relation or a relation of reduction (Horgan 1993). As such, both supervenience and realization have been used to underpin the view that mental states or properties metaphysically depend on brain states or properties but without being identical with or reducible to brain states or properties.

Central to these proprietary senses of supervenience and realization, but also to relations such as composition and constitution, is the basic assumption that has been dominant in research located at the interface between metaphysics and philosophy of mind, on the one hand, and the philosophy of cognitive science and philosophy of science, on the other, namely that mind or cognition is instantiated in the brain of individuals. This *internalist* view is usually referred to as *metaphysical*

---

[2] The concept "metaphysical building relation" is due to Bennett (2011).

[3] I borrow the distinction between "proprietary" and "generic" relations from Kaplan (2012), although he employs the distinction with regards to the boundaries and scope of mechanisms.

individualism and is (or, at least was) the default view of almost everyone in philosophy of mind and the cognitive sciences. If one takes metaphysical individualism seriously, as, for instance, the critics of EC do (see e.g., Adams & Aizawa 2008), then one also endorses the view that the object of psychological science is on the inside of the organism (see e.g., Rupert 2004, 2009). This view is commonly referred to as *methodological individualism*[4].

In philosophy, the most notorious challenge to metaphysical individualism and methodological individualism arises from Clark & Chalmers' seminal paper "The Extended Mind" (1998). They argue that mental states and cognitive processes can, in the right circumstances, occur in networks that spread across brain, body, and local environment. Against both forms of individualism, Clark and Chalmers' radical hypothesis builds an *active metaphysical externalism*. Active externalism is different from traditional meaning externalism due to Putnam (1975) and Burge (1979), since it concerns the "active role of the environment in driving cognitive processes." (Clark & Chalmers 1998, p. 7) As Clark, for instance, states:

> "I hope to convince you of at least this: that the old puzzle, the mind-body problem, really involves a hidden third party. It is the mind-body-*scaffolding* problem. It is the problem of understanding how human thought and reason is born out of looping interactions between material brains, material bodies, and complex cultural and technological environments. We create these supportive environments, but they create us too. We exist, as the thinking things we are, only thanks to a baffling dance of brains, bodies, and cultural and technological scaffolding." (2003, p. 11; italics in original)

[4] An influential example is Fodor's "methodological solipsism" (1980). On this view, one can entirely disregard facts about the body and environment when attempting to understand how cognition works. In other words, although perceptual and motor systems are reasonable objects of study in their own right, they matter little to trying to understand cognition.

1.1. Main aims, ramifications and points of controversy

The first aim of the thesis concerns the metaphysical underpinnings of EC, while my second aim is directed at research in analytical metaphysics concerning metaphysical relations such as constitution, composition, emergence, supervenience, and realization. Consequently, this thesis will have implications *both* for the metaphysical foundation of EC *and* for research in metaphysics. See figure 1 for an overview of the two main aims and how they relate. But let me first say a little about what led me to construct the core project of the thesis. I began with the goal of furthering the projects of second- and third-wave EC (Menary 2007; Sutton 2010). That is, to analyze cases of social as well as culturally distributed cognition, while arguing for the need to move away from the standard metaphysical articulations of EC turning on parity between "the internal" and "the external". However, what I found was that if this project – fully developing second- and third-wave articulations of EC – is to succeed, and if there are deep problems with the metaphysics of the EC project in general (a point often raised by the critics of EC), then one crucial ingredient is an appropriately grounded metaphysics of EC – the first main aim of the thesis.

1.1.1. First aim of the thesis

It will be argued that it is possible to *radicalize* EC even further than it already is considered to be. Despite being groundbreaking in so many respects, most articulations of EC, I submit, suffer from a failure to pay sufficient attention to the metaphysical issues their statements and theoretical concepts involve them in. It is evident from a read through the EC literature that exponents of EC use metaphysical concepts – such as constitution, composition, realization, supervenience, and emergence – to ground their metaphysical project. For example, consider these different passages from leading EC theorists:

> "*EM* is a claim about the *composition* or *constitution* of (some) mental processes." (Rowlands 2009, p. 54; italics added)

> "What is at issue, as far as the claims about cognitive extension are concerned, is simply which bits of the world make true (by serving as the local mechanistic *supervenience* base for) certain claims about a subject's here-and-now mental states or cognitive processing." (Clark 2008, p. 118; italics added)

> "Bare causal dependency of mentality on external factors – even when that causal dependency is of the "necessary" kind […] – is simply not enough for genuine cognitive extension. What is needed is *constitutive* dependence of mentality on external factors, the sort of dependence indicated by talk of the beyond-the-skin factors themselves rightly being accorded fully paid-up cognitive status." (Wheeler 2010, p. 246; italics in original)

However, from more careful reading, it is equally clear that most defenders of EC use such metaphysical concepts without scrutinizing just what these different metaphysical building relations imply[5]. Because of this[6], the metaphysical foundation of EC retains a certain theoretical and metaphysical modesty, resulting in keeping intact some of the assumptions of the tradition(s) that the defenders of EC see themselves as breaking with.

This theoretical and metaphysical modesty normally comes about through a failure to recognize the real force or impetus of the metaphysical building relations

---

[5] We should not find it too surprising that EC theorists fail to pay sufficient attention to foundational concepts and questions in metaphysics. For example, the audiences of EC and those of metaphysics do not read one another's work that often. On the other hand, EC is a new and controversial approach in cognitive science and philosophy of mind, and is, as a result, still in a stage of maturing. Also, where I emphasize that the defenders of EC do not pay sufficient attention of the metaphysical issues their use of metaphysical concepts involve them in, other theorists argue that certain defenders of EC fail to pay sufficient attention to history, culture, and to practices (see e.g, Menary 2007, 2010b; Sutton 2010). Therefore, my specific attention to the metaphysics of EC is just one of a number of important other dimensions in which further scrutiny is required.

[6] To my knowledge, Hurley (2010) was the first to point to this lack of scrutiny in the EC debate.

that hold between temporally unfolding relata such as dynamically distributed cognitive processes, on the one hand, and temporally unfolding processes such as those looping back-and-forth between brains, bodies, and environments, on the other, from which (some) cognitive processes and systems are orchestrated. That is, when defenders of EC seek to advance the *highly original* claim that human thought and reason is built – *viz*., is realized by, supervenes on, is composed by, is constituted by, or emerges from – material parts distributed across brain, body, and world, they typically keep intact the highly *unoriginal premise* of the standard view of metaphysical building relations: that the metaphysical relation that holds between cognitive phenomena and their parts *synchronically* determine, and *non-causally* explain, what they build (Kirchhoff 2013a). Call this model of metaphysical building relations SYNCHRONIC.

According to SYNCHRONIC, if an instantiation of a higher-level entity Y is synchronic, then the relation *R* that holds between Y and Y's constituents, the Xs, is exhaustively present at a single time instant *t* or entirely present at each time slice over an interval t_1, …, t_n. For example, Michelangelo's *David* is constituted by a piece of marble, *Piece*. This case, originally described in Gibbard (1975), is a standard in the literature on material constitution in metaphysics. To say that the relation *R* that holds between *Piece* and *David* exhaustively determines the existence of *David* at a synchronic instant *t* or over each point of an interval t_1, …, t_n is to say that *David*'s existence is determined *in toto* at an instant *t*, and, therefore, does not depend for its existence on unfolding over time. This follows, I submit, since it is the received view of SYNCHRONIC that the kinds of entities under discussion are (metaphysically) *enduring* entities, which are wholly present whenever they exist (Wasserman 2004b). As Horgan says about the relation of supervenience – although he may just as well have extended his claims to cover many other building relations:

> “In philosophical contexts it is primarily used *non-temporally* to signify a metaphysical and/or conceptual determination-relation.” (1993, p. 555; italics added)[7]

So, what I call SYNCHRONIC are atemporal (durationless) relations of determination between certain Xs and a specific Y, and where the relata themselves typically are understood as enduring entities which are wholly present when and where they exist.

Alternatively to SYNCHRONIC, I develop a non-standard, generic view of metaphysical building relations according to which: if the relation *R* that holds between Y and the Xs is diachronic, then *R* can never be exhaustively present at a single instant *t* or at any single time slice over an interval $t_1, \ldots, t_n$. Call this alternative view DIACHRONIC. According to DIACHRONIC, the relata – Y and the Xs – are ineliminably *time continuous* such that the relata cannot be wholly present within some particular instant *t* or at each stage across an interval $t_1, \ldots, t_n$. If we apply the language of contemporary metaphysics, the relata in DIACHRONIC are not enduring; rather, insofar as the relata of DIACHRONIC are time continuous, these relata persist by (metaphysically) *perduring*[8].

In a nutshell, it is only by giving up SYNCHRONIC and adopting my alternative view, DIACHRONIC, that EC can properly ground its metaphysical foundation, because cognitive systems and processes are themselves perduring, time-continuous phenomena, that are made up of “tangles of feedback, feedforward, and feed-around loops: loops that promiscuously criss-cross the boundaries of brain, body, and world.” (Clark 2008, p. xxviii) Thus, by adopting DIACHRONIC, the exponents of EC will be able to break with the philosophical tradition both concerning the width of the metaphysical foundation of cognition (which they do already) and in relation to SYNCHRONIC (which they do not already do).

---

[7] Or, as Gillett in this recent paper says about composition: “[…] Compositional relations are *non-causal determination relations* that are synchronous […].” (2013, p. 9; italics in original)

[8] I deal with the enduring-perduring distinction in chapters 3 and 5.

1.1.2. Second aim of the thesis

The first aim of the thesis, then, is to radicalize the metaphysical underpinnings of EC by developing an alternative DIACHRONIC conception of metaphysical building relations. Although this may seem to focus exclusively on EC, developing DIACHRONIC ultimately speaks to wider issues than just the metaphysics of EC. Indeed, such a development has implications for research in metaphysics as well. Thus, the second aim of the thesis consists in broadening the boundaries of metaphysical theorizing about metaphysical dependence relations to include not only SYNCHRONIC but equally DIACHRONIC, because dynamical processes and dynamical systems are in need of a diachronically formulated metaphysics.

1.1.3. Third-wave of extended cognition

Sutton (2010) coined the term "third-wave EC," and proposed a variety of explanatory targets for a third-wave version of EC to pursue. As he says:

> "If there is to be a distinct third wave of [EC], it might be a deterritorialized cognitive science which deals with the propagation of deformed and reformatted representations, and which dissolves individuals into peculiar loci of coordination and coalescence among multiple structured media." (2010, p. 213)

Sutton thinks that such a third-wave of EC theorizing should be an attempt to decentralize the methodological boundaries in the cognitive sciences even further and should study how history and cultural practices are active right now in the current material, social, and technological resources that enter into hybrid distributed cognitive ecologies. I agree with this. For example, in chapter 4, I develop one way by which to ground Sutton's view that a third-wave is one that 'dissolves individuals into peculiar loci of coordination and coalescence among multiple structured media'. But, where Sutton emphasizes the importance of an explanatory/methodological pathway towards a third-wave EC, I focus on articulating some of the metaphysical issues that are central to moving towards a third-wave version of EC. I should note that these two species of third-wave EC theorizing are complementary. I provide some of the metaphysics that (arguably) must be operative in a wholly hybrid and dynamical science of extended cognition. In combination with my second aim, radicalizing the metaphysical underpinnings of EC is intended to innovatively advance discussion in this field of research by joining up issues in metaphysics (concerning supervenience, realization, material constitution, composition, and emergence) with EC. It is my hope to get audiences from these different literatures to contribute more to one another's research projects, since questions which arise in metaphysics have ramifications for questions in EC, and *vice versa*.

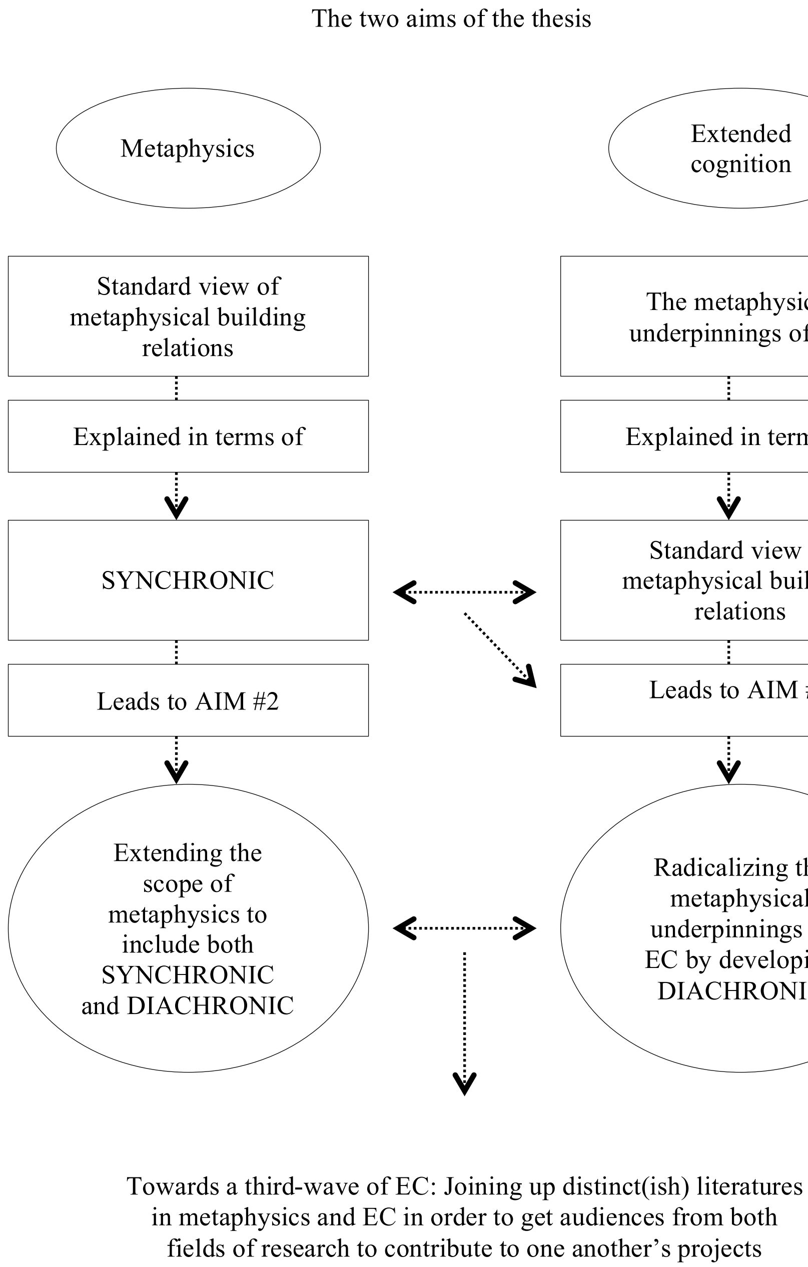

Fig. 1 Diagram of the thesis aims and how they interrelate[9]

[9] Figure 1 is intended to illustrate the focus of this thesis. It is not intended as a statement about EC in general, e.g., that EC is entirely a metaphysical thesis. Even though this thesis analyzes the metaphysics of EC, we should not forget that EC is also based closely on cognitive scientific practice, cases, and evidence. That is, EC is equally a methodological project.

1.1.4. Implications and controversy

There is a need to look carefully into just what DIACHRONIC implies, that is, its implications and points of controversy. If saying that DIACHRONIC is important due to its emphasis on relations that are diachronic – relations that unfold over time – then nearly everyone will accept DIACHRONIC, because nearly everyone, including the friends of SYNCHRONIC, give a central place to diachronic relations, specifically to causal relations. If this is all that I mean by DIACHRONIC, then that is not at all controversial, or, for that matter, original! All defenders of SYNCHRONIC include diachronic – especially causal – relations in their preferred metaphysics. As Shapiro, for example, states, when considering the relation between realization and causation: "[…], realization is intended as a synchronic relation between the realized kind and its realizer, whereas causation is most typically a diachronic relation." (2004, p. 35) Or, as Craver & Bechtel emphasize: "Mechanistically mediated effects are hybrids of constitutive and causal relations in a mechanism, where the constitutive relations are interlevel and the causal relations are exclusively intralevel." (2007, p. 547) Or, as Bennett sums up how most exponents of SYNCHRONIC think about causation and metaphysical building relations:

> "[How] *can* I avoid calling causation a building relation? There is an obvious fix: add a further necessary condition on a relation's counting as a form of building, namely that it be *synchronic*, or at least atemporal. Building relations do not unfold over time. If property *P* realizes property *Q*, it does so at some time instant *t*; if these molecules compose that table, they do so at some *t*; if these time slices compose that persisting object, they do so *simpliciter*. Causation, in contrast, is paradigmatically *diachronic*, and that idea is frequently invoked to distinguish causation from relations like composition, constitution, or supervenience […]." (2011, pp. 93-94; italics in original)

Given these typical ways of integrating diachronic relations, i.e., causal relations, into one's metaphysics, it is not the importance that I place on diachronic relations that is controversial, since, as we have seen, everyone agrees that diachronic relations (e.g., causal relations) are important. However, what is controversial about DIACHRONIC is the exclusive importance that I give diachronic relations over and in place of synchronic relations, on the one hand, and the fact that I do not restrict diachronic relations to causal relations, on the other. The most controversial aspect of DIACHRONIC is that metaphysical building relations such as *composition* and *constitution* fall under the hat of DIACHRONIC. That is a controversial claim, since almost everyone agrees that if there is anything that separates causation, on the one hand, from relations such as composition, constitution, etc., on the other, then it is the fact that the former is diachronic, whereas the latter are synchronic.

No doubt some readers are eager to offer a skeptical remark concerning the fact that I use the term "building relation" to describe diachronic relations given the intended restriction of the term "building relation" to relations such as composition, constitution, supervenience, realization, and so on. The skeptical remark might go as follows: the metaphors of "building" or "building up from", etc., seem illegitimately to smuggle diachronic phenomena to the discussion of what is usually treated as synchronic phenomena. It is the received view that metaphysical relations like composition, constitution, supervenience, and so on, synchronically determine what they give rise to. That is, if X (or the Xs) composes Y, for example, the relation that holds between X (or the Xs) and Y is understood to be a synchronic (durationless and noncausal) relation. The skeptical remark is that because "building" is typically conceived as the 'process of developing or creating something', this conception smuggles diachronic phenomena into the discussion about relations that are normally treated as synchronic relations.

But, I submit, the term *building* in metaphysical building relations is supposed to be a *generic* term for the *nature* of any one-one or many-one *relation* that holds between different *relata*. It may or may not be the case that a token building relation holds synchronically or diachronically. Building relations may be synchronic (or atemporal). But, as Bennett states, "they need not be." (2011, p. 95) Suppose that the relation between a piece of marble and a token statue is one of constitution. Constitution theorists typically consider this relation to be one that holds synchronically, fully in line with SYNCHRONIC (Wasserman 2004b). But nothing about the conception of a metaphysical building relation, I submit, necessitates that such a relation confines itself to either SYNCHRONIC or DIACHRONIC.

To be sure, there are many philosophers (see e.g., Baker 2000; Polger 2004; Shoemaker 2007; Shapiro 2004; Wilson 2004b, 2009), who treat building relations as SYNCHRONIC. But this treatment represents a commitment of certain philosophers to analyze metaphysical building relations as synchronic relations between certain entities. Such a commitment, I will argue, is not necessary, and is not entailed by the concept of a metaphysical building relation. Without assuming that metaphysical building relations must be exhaustively SYNCHRONIC, I argue that there are fruitful grounds for DIACHRONIC.

Even though my answer to the skeptical remark above might satisfy some readers, one might wonder if it is not a bit odd for a philosopher to investigate *building relations*. As Bennett puts this issue:

> "Most people who want to know how an aluminum atom is built out of various subatomic particles are asking a question for chemists or physicists. Most people, that is, are interested in the sort of building questions that are addressed by scientists, do-it-yourself manuals, and cookbooks." (Forthcoming, p. 6)

There is obviously nothing wrong with this mode of inquiry. One should indeed consult an engineer if one wishes to know how an airplane or a train or an engine is built or put together. However, this is not the kind of project that I (nor Bennett) wish to undertake by using the concept "building" with respect to metaphysical relations of determination. With regards to *metaphysical* building relations, the term "building" is intended to capture something general and abstract about the *nature* of building relations across an unrestricted set of different phenomena or of the same phenomenon. Among the categories that we will come across, there are metaphysical categories such as process, entity, time, cause, part, etc. Thus, a metaphysical statement is not intended to be a statement about everything but rather such a statement will often track one or several metaphysical categories and attempt to say something quite general and abstract about those categories. For instance, is the relationship between cognition and its physical substrate one of supervenience, realization, composition, emergence, etc.? That is the kind of question that I wish to address in the context of EC.

Am I the only philosopher investigating DIACHRONIC? To my knowledge, Bennett (2011; forthcoming) is the first philosopher to explicitly plump for diachronic metaphysical building relations. However, Bennett's aim in her forthcoming book is to establish the claim that causation is a metaphysical building relation and therefore must be included into the family of building relations. Because my aim includes establishing that composition and constitution as well as emergence are (a) all metaphysical building relations, and (b) can be reformulated in terms of DIACHRONIC, I take this project in a different direction than Bennett does. There are philosophers that advance diachronic species of emergence (see e.g., Campbell & Bickhard 2011; O'Connor 2000; Silberstein 2012). Where I depart from these authors is not in terms of my support of diachronic modes of emergence but rather in my deployment of diachronic emergence.

I will have much more to say about these issues in the rest of the thesis. I now turn to look at some of the background literature across distributed and extended cognition as well as neighboring accounts of cognition in order to position the thesis within this broader literature.

1.2. Background literature

As the reader will be aware by now, one part of the thesis concerns EC, the central tenet of which is that at least some instances of cognition arise from the productive collision points between neural and extra-neural resources. Yet there are different versions, articulating different methodologies and research interests in EC as well as neighboring accounts of cognition such as distributed cognition, dynamical systems approaches to cognition, and enactivism. To get a fix on where in this landscape the current treatment positions itself, let us take a rough-and-ready stroll through some of the defining aims and features of extended and distributed as well as dynamical and enactive cognitive frameworks.

1.2.1. Distributed cognition

The cognitive anthropologist Hutchins and his colleagues at the University of California, San Diego, have been – along with several central historical precursors such as the Soviet developmental psychologist Vygotsky (1978) – the primary forces behind the development of "distributed cognition" (DC) both as a theoretical approach and as an ethnographic methodology by which to study cognition outside the controlled set-up conditions of laboratory experiments.

In their agenda-setting paper, Hollan, Hutchins & Kirsh (2000) state that DC, just like any other theory of cognition, seeks to understand the organization of cognitive systems. Furthermore, DC, Hollan et al. mention, may be defined in accordance with the following two principles. The first principle concerns the boundaries of the *unit of analysis* for cognition. In DC, the unit of analysis pertains to the system or species of interaction that must be analyzed so as to achieve an adequate understanding of how individuals or groups of individuals cognize and behave. Sometimes these boundaries will be set by the boundaries of the individual organism, and other times the unit of analysis might include members of a social group, interactions between individuals and socially embedded as well as culturally mediated technologies, or may need to include elements distributed through time such that products of earlier events can be understood as transforming the character of later events (2000, p. 176). That is, the DC research program "looks for cognitive processes, wherever they may occur, on the basis of the functional relationships of elements that participate together in the process." (2000, p. 175) In his (1995) entitled "*Cognition in the Wild*", Hutchins published a striking exposition of DC, bringing DC's unit of analysis to encompass social interactions, human-computer interactions, propagation of information over various media, onboard a US Navy frigate.

The second principle broadens what has been the most influential view of cognition in the second half of the 20th century, namely the view that cognition is the process of manipulating symbols taking place inside the brains of individuals to include "a broader class of cognitive events and does not expect all such events to be encompassed by the skin and skull of an individual" (Hollan et al. 2000, p. 176). Consequently, from the perspective of DC, the members of a navigation team together with their tools and particular social organization make up a *cognitive system* that enables that system to accomplish such higher-level cognitive tasks as navigating a frigate safely into harbor (Hutchins 1995).

Despite its innovative widening of the unit of analysis in cognitive science, it is important to mention that while DC departs from traditional cognitive science in substantial ways, DC retains the computational model of the mind of traditional cognitive science. That is, and according to Hutchins, computation is "realized through the creation, transformation, and propagation of representational states." (1995, p. 49), and this definition can be applied to what happens both inside and outside the heads of individuals (Hutchins 1995, pp. 154-55). Consequently, DC, as a research program, takes systems larger than an individual to be a computational system, and argues that cognitive science should take, when appropriate, such larger systems as its unit of analysis.

1.2.2. Extended cognition

The DC framework was brought sharply to the attention of philosophers of mind and cognitive science when Clark adopted it as a central theoretical tool and ally in his quest to develop a rich combination and synthesis of this new species of anti-individualist movement across the cognitive sciences, resulting in Clark's publication of *Being There* in (1997) and the influential article co-written with Chalmers, "The Extended Mind" (1998).

The EC research program contains more than just one style of theorizing. There are different versions of EC with different methodologies and research interests. Sutton (2010) and Menary (2010, 2010a) have done much to make several of these important differences come to light. In the current landscape, the consensus seems to be that there are at least two articulated *waves* or *versions* of EC theorizing. Sutton (2010) calls these *first-wave EC* and *second-wave EC*, respectively, whereas Menary (2010a) distinguishes between what he calls *extended-mind-style arguments*, on the one hand, and *cognitive-integration-style arguments*, on the other. Since both Sutton and Menary intend roughly the same thing with these different terms, I stick with the distinction between first- and second-wave EC here.

First-wave EC is based on what is known as the parity principle (PP) and functionalism (Clark & Chalmers 1998; Wheeler 2010). The PP is the following statement: "If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process." (Clark and Chalmers 1998, p. 8; italics in original) The PP, as Clark stresses is a plea "for equality of opportunity" (Clark 2011, p. 451) between inner and outer processes and/or states. If the functional analogues of certain external states and/or processes were internally located, and if we would accept that these internally located states and/or processes were cognitive, then disqualifying the outer processes and/or states from being cognitive would be "philosophically unmotivated […]". (Clark 2011, p. 450) In support of the PP, Clark and Chalmers developed (amongst other cases) the example of the neurobiologically impaired Otto and his notebook. Briefly, Otto suffers from a mild form of Alzheimer's disease. Over time, Otto has written down useful information in his notebook in an apparently similar way to storing information in biological memory. According to Clark and Chalmers, because the dispositional information in Otto's notebook is functionally poised to guide action in a way that is functionally similar to non-occurrent beliefs in biological memory, the information in Otto's notebook should be considered as cognitive belief-like states. Although the PP has been interpreted as a principle of demarcation (see e.g., Adams & Aizawa 2008; Rupert 2009), Clark & Chalmers intended the PP to play the role of a methodological heuristic meant to bracket "biochauvinistic prejudice" (Clark 2008, p. 77).

Second-wave EC is based on a principle of complementary (Sutton 2010) and cognitive integration (Menary 2007), and is critical of arguments for EC based on the PP. Generally second-wave EC is EC in its integrationist, historical, and cognition-in-the-wild mode. Even so, second-wave EC is best understood not as a substantial doctrinal departure from first-wave EC; but rather, as a refinement and attunement to a more empirically and enactive-dynamical oriented approach to EC (Menary 2009). However, even if the two waves are compatible, they are also distinct, in the sense that most proponents of second-wave EC argue that the PP is either wrong or incomplete as a motivation for EC (Kirchhoff 2013b; Menary 2010b; Sutton 2010). Within second-wave EC, there are two approaches with slightly different views or inflections: the first starts from a principle of complementarity (Sutton 2010), whereas the second focuses on integration and manipulation (Menary 2007).

Sutton builds his case for complementarity by arguing, among other things, that the PP "does not encourage attention to the distinct features of the components in particular cognitive systems […]," and because of this "downplays—or even collapses—differences between inner and outer resources […]." (2010, p. 198) As a result, the PP fails to capture the dissimilarities between "inner" and "outer" parts and how they complement and "operate together in driving more-or-less intelligent thought and action." (Sutton et al. 2010, p. 525) Against functional similarity, complementarity-driven EC both predicts and requires such disparate but complementary processes between the brain's unique mode of processing and socio-culturally engineered and mediated environments.

For Menary, complementary is also an important aspect of cognitive integration (2006, p. 330). Sharing Sutton's critical stance towards the PP, Menary builds his case for second-wave EC on the manipulation thesis: "The manipulation thesis as a constituent thesis of cognitive integration is first understood to be an embodied engagement with the world, […]. Secondly it is not simply a causal relation, bodily manipulations are also normative—they are embodied practices developed through habit and training and governed by cognitive norms." (2007, p. 84) Important for Menary's version of second-wave EC is that some cognitive processes are (partly) made up of – constituted or composed by – an individual's bodily manipulation of "external" structures, with these manipulations embedded in the wider social, semantic, and normative cognitive niche (2010c, p. 611). An important focus of Menary-style EC is the idea of cognitive transformation. In particular, Menary thinks that the PP fails to explain how bodily manipulations alter the informational and physical structure of the cognitive niche, thereby transforming human cognitive capacities. Also, according to Menary, nothing in the PP tells us about just how such manipulations result in the transformation of body schemas (Gallagher 2005) required for manipulation of environmental resources, in the transformation of representational and other cognitive capacities (2010b, p. 561; see also Kirchhoff 2012).

1.2.3. Problems leveled at first-wave EC: Parity, the Martian intuition, and the grain problem

First-wave EC has been heavily attacked both by defenders of EC and by its critics. My treatment of first-wave EC would be incomplete without an exposition of at least some of these critical arguments. I start by exposing Wheeler's argument for first-wave EC. This argument turns on the PP and it is intended to show that EC is possible:

1. “If psychological phenomena are constituted by their causal-functional role, then our terms for mental states, mental processes, and so on pick out equivalence classes of different material substrates, any one of which might in principle realize the type-identified state or process in question.” (Wheeler 2010, p. 248)
2. “If there is functional equality with respect to governing behavior, between the causal contribution of certain internal elements and the causal contribution of certain external elements, and if the internal elements concerned qualify as the proper parts of a cognitive trait, then there is no good reason to deny equivalent status - that is, cognitive status - to the relevant external elements.” (Wheeler 2010, p. 248).
3. If parity of causal contribution mandates parity of status, and if mental states and processes are multiply realizable, then “it is possible for the very same type-identified cognitive state or process to be available in two different generic formats - one non-extended and one extended.” (Wheeler 2010, p. 248)
4. *Therefore*: Cognitive processes and states are realizable (partly, at least) by external states and processes (Wheeler 2010, p. 249).

Why disagree with this? After all, if, and whenever, two type-identified states (or processes) play the same causal-functional role, it is (in principle) irrelevant whether neural or non-neural elements realize that role.

The first reason for being skeptical about parity-driven arguments for EC, we are familiar with, namely that the requirements of parity downplay – or even collapse – some important differences between internal and external elements (Sutton 2010, p. 199). According to Haugeland: “Such arguments are indifferent to variety and substructure within either the mental or the physical: everything is unceremoniously lumped together at one swoop.” (1998, p. 228) Thus, the first problem with PP arguments is that they pay insufficient attention to differences between inner and outer resources.

Another influential critique of first-wave EC is the *Martian Intuition* argument raised by Sprevak (2009). This argument turns on the fact that first-wave EC is based on functionalism. Sprevak starts by acknowledging that functionalism was engineered, in part, so as to save the Martian Intuition: that it should be possible for creatures with cognitive processes to exist even if such creatures have a different physical and biological makeup from human beings (Block 1980). According to Sprevak, if defenders of EC employ functionalism to ground EC, then the defender of EC faces the following dilemma: (a) accept functionalism and radical EC; or, (b) give up EC entirely (2009, p. 503). Sprevak thinks that if one accepts functionalism (which grounds the PP), then it entails a commitment to a wildly implausible or too radical version of EC, the consequence being: "rampant expansion of the mind into the world […]." (2009, p. 503) Here is the form of Sprevak's argument: (1) Functionalism entails the Martian intuition (P); (2) if P, then radical EC (Q); (3) P is true; and (4), therefore, Q is true.

So, if functionalism entails the Martian intuition, and if functional parity is used to save or ground the metaphysical claim of EC, it follows that the defenders of parity-based EC must be committed to a radically implausible view of cognition, where the following scenario holds: simply by picking up a book, one comes to believe all the information contained in that book. The justification for this claim is: (a) a Martian might "internally" encode memories in ink-marks; (b) in addition to gaining its beliefs via sense modalities such a Martian might equally be born with innate beliefs; (c) moreover, it is possible that the Martian might have such innate beliefs that it has not yet examined, *viz*., that the Martian has a library of data phylogenetically hardwired into its cognitive system; and (d) finally, it is possible to imagine that this Martian has such a stock of innate beliefs stored in an ink-based memory system, most of which it has not yet had any reason (or cause) to employ. Sprevak's point is that it is quite plausible to think that such a creature could exist. As Sprevak says, the:

> "Martian has ink-marks inside its head that, if it were sufficiently diligent, would guide its action in appropriate ways. The difference between the Martian and me is that it has the ink-marks inside its head, while I have the ink-marks outside. By the fair-treatment principle [the PP], if the Martian has beliefs, then so do I." (2009, p. 518)

Even if Sprevak is correct in his assessment, matters are more complicated. Within first-wave EC, Wheeler, in order to circumvent Sprevak's conclusion, has suggested the following move. One should attempt to go in between the horns of the dilemma by arguing that the relevant level of grain by which we individuate functional roles should be set neither too high (so as not to entail radical EC) nor too low (so as to block the critics' difference-argument)[10]. But, according to Rowlands, attempting to establish the relevant level of granularity with respect to functional roles leads to an impasse. That is, a deadlock, the ramifications of which will have paralyzing effects on functionalist arguments, pro and con, for EC. As Rowlands states:

> "If Rupert's arguments against the extended mind are question-begging because they presuppose a chauvinistic form of functionalism, it is difficult to see why arguments for the extended mind are not question-begging given their predication on a liberal form of functionalism." (Rowlands, unpublished ms, pp. 6-7; quoted in Wheeler 2010, p. 255; see also Rowlands 2010, pp. 209-10)

In what we might call Rupert-style anti-EC (2004, 2009), the most common way of criticizing the PP consists in noting a set of psychological properties found in human neural-cognitive systems but not socio-cultural systems, and then inferring that there is no parity at the level of fine-grained functional operations between "inner" and "outer". Hence, inferring the metaphysical claim of EC from parity-based arguments must be false, on Rupert's view. The defenders of parity-based EC commonly respond to this line of argument by charging the Rupert-style argument of advocating a chauvinistic form of functionalism. As Wheeler says: "[…] it seems that Rupert's […] argument continues to beg the question against extended functionalism […], extended functionalism looks to be predicated on the more liberal form of functionalism that generates a locationally uncommitted account of the cognitive." (2010, p. 225) It is from this problem – the problem of identifying the appropriate level of functional grain – that Rowlands concludes that the debate

[10] The critical "differences argument" (Adams & Aizawa 2001; Rupert 2004, 2009), starts with the observation that there is an obvious distinctiveness between the *fine-grained* causal-functional profile of "internal" operations and the causal-functional profile of "external" operations, and infers from this that because functional parity arguments hold that there is functional similarity between "inner" and "outer", such arguments fail to establish the possibility of EC.

over constitution by way of functional parity has fallen into an impasse, with both sides potentially begging the question against one another.

With these different criticisms of first-wave EC mentioned, the next couple of accounts of cognition on our list are dynamical systems and enactivism.

1.2.4. Dynamical systems & enactive approaches to cognition

The revolt against traditional scientific cognitive methodologies and theoretical assumptions – the view that cognition is computation over symbolic representations instantiated in the brain – finds its most radical articulation in approaches to cognition inspired by insights from dynamical systems theory (see e.g., Chemero 2000, 2009; Chemero & Silberstein 2008a, 2008b; Keijzer 2001; Port & van Gelder 1995; van Gelder 1995, 1998) and in the enactivist framework (see e.g., Di Paolo 2009; Hutto & Myin 2013). I am aware that there are differences both within each camp as well as between the two approaches. But for my purposes here, the fact that there are significant overlap in methodology and theoretical assumptions is good enough. For instance, most enactivists use dynamical systems theory to explain their experimental results, from neuroscience (Di Paolo et al 2008; Froese & Di Paolo 2010), social modes of cognition (Di Paolo & De Jaegher 2012; De Jaegher & Di Paolo 2013), to debates about the life-mind continuity thesis (Thompson 2007). Thus, for simplicity, I restrict my exposition to dynamical-systems-theoretic accounts of cognition. Consider the following quote by van Gelder & Port:

> "The heart of the problem is *time*. *Cognitive processes and their context unfold continuously and simultaneously in real time.* Computational models specify a discrete sequence of static internal states in arbitrary "step" time (t1, t2, etc.). Imposing the latter onto the former is like wearing shoes on your hands. You can do it, but gloves fit a whole lot better." (1995, p. 2; italics in original)

From this observation – that temporal unfolding in a dynamical context is inherently part of the *nature* of cognitive processes and modes of processing – the dynamical systems theorist typically infers the following claim: that conceiving of cognitive processing as a form of computational processing ignores real time. The claim that the phenomenon "computation" ignores real time is meant to show that understanding cognitive processing as computational inherently ignores the fact that dynamical systems – which the dynamicists take cognitive systems to be – are quantitatively embedded in time. As Smithers points out: "If we change this embedding by slowing down all the movements and actions, or by speeding everything up, we change the behavior, and it will no longer be of much use. Intelligent behavior thus has an important and essential underlying dynamic […]." (1998, p. 652) In contrast to dynamical systems, the claim is that computation downplays the role of temporal unfolding. That is, as Smithers states: "Changing the rate at which the computation is done makes no difference; it remains the same computation." (1998, p. 652; see also van Gelder 1998, p. 618) If this is correct, as the defenders of the dynamical approach to cognition insist it is (see also Chemero 2009; Spivey 2007), and if cognitive systems are dynamical systems, then the dynamicist infers the radical conclusion that the computational theory of mind must go.

That concludes my rough-and-ready survey of views in philosophy and cognitive science, aiming to explain how cognition is multiply distributed. Of course, the data and theoretical arguments that may be said to impress these theorists go well beyond what is surveyed in the previous sections. However, it will be enough to allow me to situate the main aims and contributions of the thesis within the current literature.

1.3. Positioning the thesis in the literature

This thesis has two distinct (but complementary) aims. The first is to radicalize the metaphysics of EC. The second is to develop DIACHRONIC. I will first say something about how these two aims relate.

1.3.1. How my two aims relate

It is entirely reasonable to accept DIACHRONIC, but still reject EC. However, I submit, if one wants to defend EC on metaphysical grounds, it is not possible to reject DIACHRONIC.

First, much of the urge towards EC derives from considerations about "densely coupled unfolding" (Clark 2005b, p. 234) in stables of dynamical and embodied cognitive science. Because EC is often based on considerations about dynamical systems, it would seem that EC must equally take seriously the idea that cognitive processes and their context unfold continuously in real time, and that time cannot be reduced to a set of discrete quanta such as t1, t2, t3, etc. And indeed they do. For instance, in Clark's (1998b) article "Time and Mind," Clark agrees with van Gelder and Port that cognitive processes are *continuous processes*. Clark gives this explanation of what a continuous process is: "A continuous process is one in which the time-series of explanatorily relevant sub-states cannot be reduced to a sequence of discrete states with jumps in between, but instead requires a genuine continuum of states." (1998b, p. 357) Or, as Wheeler states, when he says that Turing machine computation is *temporally impoverished*: "the system features a style of processing in which time is reduced to mere sequence." (2005, p. 105)

Here is the culprit: If distributed cognitive processes, and the dynamical systems instantiating distributed cognitive processes, unfold continuously in time, then the relationship between a distributed cognitive process, say, and the densely coupled parts giving rise to that distributed cognitive process, can never hold completely at a specific *synchronic* time instant *t* but must hold *diachronically*. In other words, because the standard view of metaphysical building relations reduces time to discrete quanta (SYNCHRONIC), and because dynamically distributed processes are time continuous, SYNCHRONIC is inconsistent with distributed cognitive processes.

Another core feature of my critique of the received, synchronic view is that the sort of entities presupposed as the relata of synchronic building relations are enduring entities which are wholly present whenever they exist. But, insofar as processes in general and cognitive processes in particular cannot be wholly present at a specific synchronic time instant, it follows that such processes cannot be understood as enduring entities. Thus, to adequately ground its own metaphysical foundation, EC must go DIACHRONIC. DIACHRONIC requires that we rethink the metaphysics of building relations and the metaphysical underpinnings of EC.

One can, as I mentioned, plump for my diachronic account without taking EC onboard. This is why this thesis ultimately appeals to wider issues than merely those in EC. For example, both constitution and composition are commonly expressed in terms of SYNCHRONIC. A paradigmatic case is the relationship between *David* and *Piece*. However, insofar as either the constitution relation or the composition relation (or, both) may hold in dynamical systems – regardless of those systems being cognitive – SYNCHRONIC is ill equipped to analyze the relevant relation of dependence between the whole and its parts. Because DIACHRONIC is both a general framework and applicable to the subject matter of EC, and because distributed cognitive processes are instantiated in distributed dynamical systems, DIACHRONIC, but not SYNCHRONIC, provides an adequate set of conceptual tools to make sense of dependence relations in dynamical systems. Note, though, because DIACHRONIC is intended to be a generic framework, it follows that it is possible to accept the view of metaphysical dependence relations put forth in DIACHRONIC, yet deny EC (on other grounds). So, there is no entailment relation from DIACHRONIC to EC.

1.3.2. How the thesis sits in the literature

I take from the dynamical and enactive approaches the idea that temporal unfolding is an ineliminable part of cognitive activity. It is a central tactic of the later chapters to establish that insofar as the metaphysics of EC is concerned, downplaying the dynamics of time should be avoided. This brings my metaphysical project into alignment with core tenets from both dynamical and enactive cognitive science. However, I only take from dynamical and enactive cognitive science the idea that cognitive systems and cognitive processes are time continuous, together with central concepts such as self-organization and nonlinearity[11]. I am not dispensing with computation and representation for the simple reason that I do not think that the very notions of computation and representation are "intrinsically" temporally austere[12].

---

[11] My interest here is in physical time, not the psychology of time (or, time perception). An intriguing investigation would integrate issues pertaining to time perception. Unfortunately I do not have space and time to do this – a task for another occasion.

[12] In chapter 6, where I look at the relationship between the realization relation and the free energy minimization formulation in cognitive neuroscience, we will come

Instead of doubting the concepts of computation and representation, I explore this question: If we take seriously the idea that cognitive processes and their contexts continuously unfold in real time, why, then, do philosophers of mind and cognitive science keep appealing to relations such as supervenience and realization that are considered to hold synchronically between their relata? In other words, how can a synchronic relation hold between exclusively diachronic (i.e., temporally and causally unfolding) processes? I explore, and will argue for, this deep incompatibility in chapters 3, 4, 5, 6, 7, and 8.

This brings us to EC. There are independent reasons for being skeptical about first-wave EC. That is, issues regarding the PP, the Martian intuition, and the grain problem. Although I agree that these are all serious problems for first-wave EC, they are not the ones that I wish to emphasize (see Kirchhoff 2013b for discussion of these three problems with first-wave EC). My beef is with the failure of defenders of first-wave EC to pay sufficient attention to the metaphysical issues that their use of such relations as supervenience and realization, for example, involve them in.

What about second-wave EC? The alternative DIACHRONIC framework that I develop in the thesis is not dependent on either the complementarity principle (Sutton 2010) or cognitive integration (Menary 2007). So, it is not part of my project to engage in a discussion about whether or not particular second-wave arguments are better than particular first-wave arguments. But, DIACHRONIC, I submit, is implicit in non-functionalist accounts of second-wave EC, on the one hand, and in dynamical and enactive cognitive science, on the other, even if DIACHRONIC views have not been explicitly stated, and even though – as will become clear throughout this thesis – these DIACHRONIC views undermine other views that are commonly endorsed[13]. Therefore, I attempt to further the project of second-wave EC, albeit I do so by articulating a diachronic metaphysics of

---

across notions of both computation and representation, which are entirely compatible with richly temporal processes and modes of processing.

[13] For instance, when I discuss the causal-constitutive fallacy, I analyze one particular example by Menary (2006) that involves the constitution of a distributed process of remembering, and I show that this notion of constitution is wholly consistent with a diachronic interpretation.

dependence relations that moves beyond the typical ambitions of second-wave EC and towards a third-wave version of EC.

1.4. Strategy and methods

In an attempt to radicalize the metaphysical foundation of EC, on the one hand, and in providing an alternative DIACHRONIC conception of metaphysical building relations, on the other, I use the novel distinction between *ontological synchronicity* (roughly, at a particular time instant *t*) and *ontological diachronicity* (roughly, over an interval of time). Utilizing this distinction allows me to specify the following: metaphysically speaking, the synchronic timescale and the diachronic timescale are mutually exclusive such that it cannot be true of one and the same phenomenon – at one and the same level of analysis – that the building relation in question is both ontologically synchronic and ontologically diachronic. Throughout the thesis, I shall make use of this distinction to highlight the limitations of ontological synchronicity concerning metaphysical building relations in dynamical systems.

A note on the use of case studies in relation to DIACHRONIC throughout the thesis. As I mentioned, the diachronic framework is a generic framework, one that is not restricted to the subject matter of EC, in that, it applies to dynamical systems in general, irrespective of these being cognitive systems. Thus, throughout the thesis, I shift between considering cognitive systems and non-cognitive systems. But, when I do consider cognitive systems or cognitive processes, I pick the particular case studies that I do, because I want to emphasize the socio-cultural rather than the strictly artifactual dimensions of research in EC, thereby nudging the projects of second- and third-wave EC along further (see e.g., Kirchhoff & Newsome 2012; Menary 2010b; Sutton 2010).

1.5. Overview of the thesis chapters

The plan of the thesis is as follows. Chapter 2 is expository. It is my aim to survey most of the familiar metaphysical building relations in order to establish that these relations are standardly thought to express SYNCHRONIC. In addition to this, I provide a survey of the concept of "metaphysical building relation".

In chapter 3, I take a few steps towards DIACHRONIC, before applying this view to issues pertinent to EC in the rest of the thesis. I start by establishing, in chapter 3, that SYNCHRONIC can be criticized by looking at few examples of dynamical systems.

In chapter 4, I discuss a recent divergence between Clark (2008, 2011) and Hutchins (2011a) concerning the timescales over which processes combine to jointly assemble instances of extended cognitive processes and/or systems. I argue that even though Clark favors the short-term timescales of the here-and-now, whereas Hutchins targets slower timescales, both approaches are incompatible with SYNCHRONIC.

In chapter 5, I show that DIACHRONIC – but not SYNCHRONIC – has the apparatus required for analyzing cases of socially distributed cognition. I base this argument on an analysis of transactive remembering.

In chapter 6, I address the realization relation. In particular, I test the flat view and the dimensioned view (Gillett 2002, 2007a), on the one hand, and the wide view (Wilson 2001, 2004a, 2004b), on the other, against the free energy principle in cognitive neuroscience (Friston 2010). I do this for three reasons. The first is that the free energy principle has recently been argued to be a powerful ally to EC (Clark 2013). The second reason is that the free energy principle portrays the mind-body relationship as one of free energy minimization, and the most prominent application of the realization relation has been to address the mind-body problem. As a first approximation, this suggests that the realization relation and free energy minimization can be brought together. My final reason is that in various writings, Wilson (2001, 2004a, 2004b) has argued that if the relation of realization is metaphysically wide, the realization relation may be used to ground the metaphysics of EC. By extension, then, if EC is consistent with the free energy principle, and if the metaphysics of EC is consistent with wide realization, then the free energy principle is consistent with wide realization.

Research at the interface between philosophy of mind and philosophy of cognitive science, on the one hand, and philosophy of science and metaphysics, on the other, is filled with claims about emergence. It is no different in EC. Hence, in chapter 7, my goal is to begin to develop an ontological diachronic notion of emergence (see e.g., Campbell & Bickhard 2011; Kirchhoff 2013c; Mitchell 2012; Silberstein & McGeever 1999), which leads to the view that the synchronic account of supervenience emergentism is problematic. Having argued that the SYNCHRONIC view of supervenience emergentism is problematic, I move on to pursue my second aim, namely to explore an implication of my argument against supervenience emergentism for EC. The upshot will be the suggestion that EC should avoid supervenience talk unconditionally.

In the final chapter, I consider what we might call a classic in the EC literature. I launch a critical argument against the alleged causal-constitutive fallacy leveled against EC. Fortunately for us, all the hard work of establishing DIACHRONIC now pays for itself. That is, I shall argue that the critics of EC – such as Adams & Aizawa – are wrong to charge EC with the universally construed causal-constitutive fallacy, since Adams & Aizawa are working with a notion of SYNCHRONIC constitution that is inconsistent with common cases of EC. Thus, paying closer attention to the metaphysical concepts we apply may provide us with all the tools we need to successfully deal with our critics.

1.6. Overview of the thesis's conclusions

This chapter has previewed the thesis' overarching themes and clarified its goals and methods. I wish to finish this introduction by clarifying some of the thesis' main conclusions. Consider, first, figure 2, which lists five different SYNCHRONIC building relations:

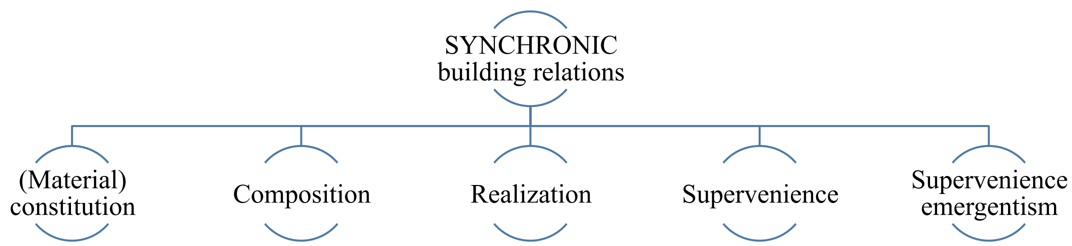

Fig. 2 Overview of five different synchronic building relations

Based on issues such as decentralization, self-organization, nonlinearity, context-dependence, relational properties, dynamical systems, and processes, I argue (chapters 4, 5, 6, 7, and 8) that the metaphysical foundation of EC can be grounded by appeal to constitution, composition, and emergence only if none of these building relations are understood in the terms of SYNCHRONIC. Thus, in those chapters, I defend and reformulate the following three building relations in terms of DIACHRONIC (figure 3):

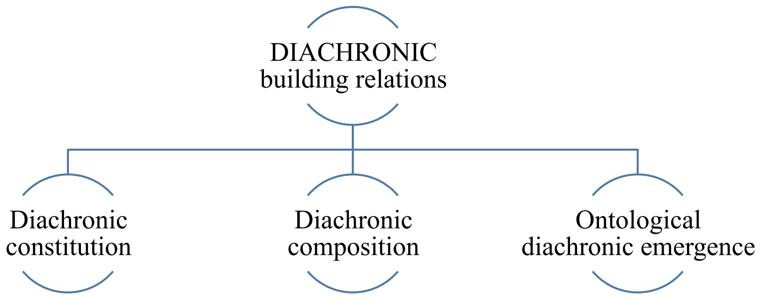

Fig. 3 Overview of three different diachronic building relations

I problematize both realization and supervenience as these are currently construed as candidates for the project of underpinning EC's metaphysical foundation. That is, I argue, in chapters 6 and 7, that realization and supervenience carry with them baggage that should not be accepted by the EC theorists – e.g., appeals to ontological synchronicity and intrinsic properties. Given the current state of evidence from dynamical systems, I suggest that we should abandon both of these features from our metaphysics – at least when addressing systems and/or processes grounded in self-organizing, nonlinear, temporal and decentralized dynamics. Thus, I arrive at the conclusion that insofar as the metaphysical foundation of EC is concerned such an endeavor must go completely DIACHRONIC. In addition, related to my second aim of the thesis, I conclude that plumping for DIACHRONIC allows research in metaphysics to engage with cases of metaphysical building relations far more dynamic and complex than many of the atemporal examples commonly used in the metaphysical literature.

2. Metaphysical building relations: A survey of synchronic relations

This chapter will be mostly expository. I will focus on surveying metaphysical building relations that express what I labeled the SYNCHRONIC account in the introduction. Together with introducing a general framework for thinking about the term "metaphysical building relation," what I hope to show here is that the standard view of metaphysical building relations presuppose SYNCHRONIC.

Let us start with the observation that one theme that occupies a large space throughout philosophy is what I, following Bennett (2011), call *metaphysical building relations*. A read through such fields as metaphysics, philosophy of cognitive science, philosophy of mind, philosophy of science, etc., reveals much talk of something's being built up from, giving rise to, getting out of, determining, putting together, assembling, etc. For example, in the philosophy of mind, one finds questions about how mental properties are instantiated by or dependent on non-mental, physical properties. In metaphysics, one question that is sometimes asked is how nonmodal properties give rise to modal properties. In philosophy of science, where talk of levels is commonplace, one often finds questions pertaining to how macro-level properties are constructed from micro-level properties. In ethics, some philosophers have been occupied with the question of how moral properties get out of non-moral properties. And so on.

Metaphysical building relations are the relations that philosophers commonly use to come up with answers to questions such as those mentioned above by invoking relations such as supervenience, realization, emergence, constitution, composition, and others.

2.1. Aim and overview

In the philosophy of extended cognition, the area of research that will occupy me in the thesis, the core metaphysical building question is: How are cognitive processes or systems *built up from* physical processes or systems crisscrossing such heterogeneous elements as the brain, body, and environment? Or, under what circumstances are certain cognitive processes or systems grounded metaphysically in elements distributed across brain, body, and world? As one of the leading exponents of EC says: "Thus if the extended [cognition] thesis is true, it is true in virtue of something *implementationally deep* about cognition." (Wilson 2010, p. 171; italics added)

Often this metaphysical question is phrased in terms of *constitution* in the EC literature. Relations such as constitution fall under what I, following Bennett (2011), call a metaphysical building relation. In discussing the metaphysical underpinnings of EC, I keep in use the notion of a metaphysical building relation for the reason that it is not only the constitution relation that is invoked by defenders of EC but also a variety of other metaphysical relations such as supervenience, composition, realization, and emergence.

The overarching aim of this chapter is expository. I wish to survey five different but very familiar metaphysical building relations, all standardly understood to express SYNCHRONIC. I will begin by characterizing the central features of the concept "metaphysical building relation". Then I survey five different building relations in section 2.3.

2.2. Metaphysical building relations

A useful starting point in an analysis of what characterizes a metaphysical building relation is the idea that for *R* to qualify as a metaphysical building relation, *R* must express the form 'X (or the Xs) metaphysically determines Y', when it is *in virtue of* X (or the Xs) that Y exists. This *in virtue-ness* is often specified as a species of determination (see e.g., Kim 1990; Polger 2010; Shapiro 2004). That is, if X (or the Xs) builds Y, metaphysically, then X (or the Xs) determines the existence of Y. Different relations – such as constitution, composition, realization, supervenience, emergence, and others – have been used in philosophy to express the view that something exists in virtue of something else (Bennett 2011; Kim 1998).

From the "standard view," as I call it, of a metaphysical building relation, it is commonplace to say of this kind of determination that it, at least, holds *irreflexively* (i.e., if X (or the Xs) determines the existence of Y, then neither X (or the Xs) nor Y determines the existence of themselves) and *asymmetrically* (i.e., if X (or the Xs) determines the existence of Y, Y does not determine X (or the Xs) on the occasions when a relation holding counts as building). Following Bennett, we can highlight two things about the fact that metaphysical building relations hold asymmetrically and irreflexively. First, building relations are relations of *relative fundamentality* (RF). Formally: "(RF) for all *x* and *y*, and all building relations *B*, if *xBy* then *x* is more fundamental than *y*." (Bennett, forthcoming, p. 27) If X (or the Xs) composes Y, or if Y is emergent from X (or the Xs), in both cases Y is *in some sense* less fundamental than its base, X (or the Xs)[14].

That building relations have this implication is not surprising. It is precisely the implication expressed by philosophers who use relations such as supervenience, realization, constitution, and so on, to state that Y exists in virtue of X (or the Xs)[15]. The second thing is that "building relations have an input-output structure; they take some relatum(a) and generate another." (Bennett 2011, p. 91) That is, if *xBy*, then *B*

[14] Depending on which sense of "emergence" one is using, it is possible to claim that emergence is both a relation that expresses that emergent properties are less fundamental than their base and that emergence is not a relation of relative fundamentality. There is nothing (in principle) problematic about this so long as the sets of claims – pertaining to the different senses of emergence – do not entail a contradiction. For a view of emergence that would seem to exclude emergence from being a relation of relative fundamentality see Chalmers (2006; see also Bennett, forthcoming, p. 27). I deal extensively with emergence in chapter 7.

[15] As Kim, for example, states in his discussion of supervenience, it is customary to associate supervenience with the idea of dependence such that "if [y] depends on, or is determined by [x], it cannot be that [x] in turn depends on or is determined by [y]. What does the determining must be taken to be, in some sense, ontological prior to, or more basic than, what gets determined by it." (1998, location 247/2719; Kindle Version)

takes as its ‘input’ *x* and as its ‘output’ *y*, and insofar as *B* expresses a relation of RF, it follows that the ‘input’ of *B* is in some sense more fundamental than its ‘output’.

From this Bennett gives the following, minimal, requirement that any relation, for it to qualify as a metaphysical building relation, must imply: “[A] relation is a building relation if and only if:

- it is asymmetric and irreflexive,
- the ‘input’ relatum(a) is both more fundamental than the ‘output’, and
- Either:
  the input is minimally sufficient in the circumstances for the output, or
  the existence, instantiation, or occurrence of the output counterfactually depends on that of the input.” (Forthcoming, p. 32)[16]

Metaphysical building relations are often contrasted with the relation of identity. Identity relations are reflexive, symmetrical, and transitive (Wilson 2009, p. 363). What about transitivity? In his discussion of material constitution, Wilson says that in metaphysics, “the view that material constitution is transitive is ubiquitous, an assumption expressed by both proponents and critics of constitution views.” (2009, p. 363) Transitivity entails, if X (or the Xs) constitutes Y, and if X (or the Xs) is constituted by Z (or the Zs), then Y is constituted by Z (or the Zs). To be sure, then, there are philosophers who argue that insofar as the relation between X (or the Xs) and Y is one of relative fundamentality, then the relation between X (or the Xs) and

[16] Let me quickly deflect a potential misunderstanding concerning emergence. Even though dynamical systems are commonly said to exhibit emergent phenomena, and even though many dynamical systems exhibit both top-down and bottom-up mediated effects (Kelso 1995), this particular “symmetrical relation” between higher and lower levels does not contravene the idea that emergence is asymmetric. If Y is emergent from the Xs, the *existence* of Y is determined by the Xs. However, there is nothing problematic about the idea that Y may affect the functioning of the Xs so long as we do not confuse the two kinds of relations. I deal with this issue in chapter 7.

Y is transitive. But there is no entailment relation between relative fundamentality and transitivity. Consider, e.g., the following statement by Bennett:

> "Suppose that, for some building relation *B*, *aBb*, *bBc*, but there is no *B* such that *aBc*. (RF) entails that *a* is more fundamental that *b*, and *b* is more fundamental than *c*. The transitivity of *more fundamental than* entails that *a* is also more fundamental than *c*. But that does not conflict with the assumption that *a* does not build *c*: it simply has to be the case that one thing can be more fundamental than another despite not standing in a building relation to it. This is clearly true. Intuitively, sodium ions are more fundamental than benzene rings, but benzene rings are not even partly built out of sodium ions – sodium is simply not involved." (Forthcoming, pp. 28-29)

Another way to express the idea that there is no entailment relation between the notion of a metaphysical building relation and transitivity is due to Wilson (2009). For instance, consider these two arguments:

1.A. This chain is constituted by metal links.
1.B. Those metal links are constituted by physical particles.
1.C. This chain is constituted by physical particles.

In this case, the premises (1.A) and (1.B), together with the criterion of transitivity, entail (1.C). What about the following argument?

2.A. This queue is constituted by a sequential order of people.
2.B. That sequential order of people is constituted by physical particles.
2.C. This queue is constituted by physical particles.

Argument 2A-2C has the same form as 1A-1C. However, even if both arguments rely on the principle of transitivity, unlike 1A-1C, 2A-2C is controversial, in that, it is not clear that 2A-2C can accommodate transitivity. Specifically, unlike a metal chain, which one might think of as *nothing more than* various entities appropriately organized, queues are *more than simply* their physical parts – regardless of how these might be arranged. Hence, while it is entirely unproblematic to say of argument 1A-1C that it is transitive, the same is not true of argument 2A-2C (for a detailed discussion, see chapter 8). Consequently, I refrain from adding the formal property of transitivity as a necessary condition for a relation to count as a metaphysical building relation, since even though metaphysical building relations are relations of relative fundamentality, it does not follow that all instances of building relations are transitive.

From these minimal requirements, one further condition is standardly added for a relation to count as a metaphysical building relation, namely that it must hold *synchronically*. For instance, in his (2004) and (2011), Shapiro says this about the relations of realization and constitution, respectively:

> "As I have already mentioned in passing, realization is intended as a *synchronic relation* between the realized kind and its realizer." (2004, p. 36; italics added)

> "[If] C is a constituent of an event or process P, C exists where and when that event or process exists. Thus, for some process P, if C takes place prior to P's occurrence […], or if C takes place apart from P's occurrence […], then C is not a constituent of P." (2011, p. 160)

This added necessary condition of synchronicity is typically made by invoking a distinction between relations that hold *synchronically* and those that hold *diachronically*, thus demarcating diachronic relations from the family of metaphysical building relations. Note that this distinction is usually drawn in order to separate *causation* from counting as a metaphysical building relation. Here Shapiro is explicit once again: "Thus, the synchronic nature of realization serves to distinguish it from causation […]." (2004, p. 36). Or, as Bennett points out: "Causation, in contrast, is paradigmatically *diachronic*, and that idea is frequently invoked to distinguish causation from relations like composition, constitution, or supervenience […]." (2011, pp. 93-94; italics in original)[17]

[17] Bennett does not endorse this way of distinguishing between metaphysical building relations and diachronic relations. In fact, she immediately rejects this, as she says: "Yet although this is a tempting solution to the problem, occasionally

Philosophers accepting the condition that only relations that hold synchronically may count as metaphysical building relations often find it natural to say that metaphysical building relations are *noncausal* relations of determination[18]. For instance, in his discussion of realization, Polger emphasizes:

> "Whether in the restricted or unrestricted form, the thesis is meant to be ontological: one entity or set of entities ontologically depends on another. The dependence is *non-causal* […]." (2010, p. 195; italics added)

That a metaphysical building relation is standardly conceived of as synchronically determining (and non-causally explaining) what it constructs can easily be made sense of with the following two examples: (i) the statue in front of me is constituted by a piece of marble; and (ii) throwing a bottle on a statue in front of me causes the bottle to break.

The intuition is that only one of these two examples is a case of metaphysical building, since only example (i) implies a *vertical* relation of determination, whereas example (ii) implies a *horizontal* relation of dependence. The horizontal

---

temptation is best resisted. This is one of those occasions. We should not require that building relations be synchronic, because there is at least one important relation that is worth calling a building relation, but that unfolds over time. I explore this elsewhere (MSa)." (Bennett 2011, p. 94) As I mentioned in the introduction, Bennett's aim is to establish that causation is a metaphysical building relation. In this thesis, my primary goal when addressing the possibility of *diachronic* metaphysical building relations is not to propose that causation is a building relation but rather to establish that relations such as constitution, composition and emergence are ripe for a diachronic articulation.

[18] Note that in contrast to the relation of causation, insisting on the idea that only metaphysical building relations synchronically determine (and thus noncausally explain) what they give rise too does not exclude causation from being a relation of determination. To be sure, exponents of the idea that metaphysical building relations determine higher-level phenomena synchronically, will, all things being equal, endorse the view that causation is a diachronic relation of determination such that causes diachronically determine their effects.

dimension represents *time*, whereas the vertical dimension represents a synchronic relation of relative fundamentality (see e.g., Aizawa & Gillett 2009a). In a nutshell, if I throw a bottle at some statue in front of me, a time interval will unfold between me throwing the bottle and the bottle breaking, whereas if the statue itself is constituted by a piece of marble, the constitution relation between a piece of marble and the token statue holds instantaneously (at a durationless point in time). This is not only the case for constitution, but is usually taken to be the case for all metaphysical building relations (Bennett 2011; Gillett 2007b). Again we can find a statement by Shapiro highlighting this distinction:

> "A realization is present *simultaneously* with that which it realizes and cannot be separated from it. In contrast, causes and effects are independent events. A cause is followed by an effect and, depending on the *amount of time* between the cause and the effect, it is possible to imagine that a cause and its effect never exist simultaneously." (2004, p. 35; italics added)

That concludes my brief sketch of the standard notion of a metaphysical building relation. In summary, a metaphysical building relation is (a) a synchronic (atemporal) dependence relation; (b) a relation of relative fundamentality; and (c) asymmetric, irreflexive, and (usually) transitive. The aim of the next section is to survey five familiar building relations, sketch a few differences between them, point to how each plays its distinctive building role, and show that they all hold synchronically.

2.3. A survey of SYNCHRONIC building relations

Due to the large literature on relations like supervenience (see e.g., Horgan (1993)), and other building relations, I consider here, including all their ins and outs is not a feasible option. Instead my aim is to say just enough about each building relation to introduce these relations and explain how they differ from one another. For an overview of SYNCHRONIC building relations, see figure 2 (chapter 1). I begin this survey with the relation of material constitution.

2.3.1. Material constitution[19]

Michelangelo's *David* is constituted by a particular piece of marble, *Piece*. These metal links constitute this token metal chain. Or, this particular Danish national flag is constituted by a piece of fabric. These are familiar examples of what is called the *material constitution view* in metaphysics.

Material constitution is standardly taken to be a *synchronic* one-one relation of determination that holds between spatially and materially co-located objects of different kind. In the literature, there is still some debate concerning the claim that constitution holds between two distinct objects that exists at the same time and in the same place (see e.g., Wasserman 2009). Despite disagreement about just how objects can exist at the same time and in the same place, but still differ with regards to their modal properties, the constitution view is popular, and has been defended by Baker (1997, 2000), Chappell (1990), Fine (2003), Johnston (1992), Kripke (1971), Lowe (1995), Shoemaker (1999), Simons (1985), and others.

The constitution relation can be framed in terms of how to fill out the following schema: X (or the Xs) constitutes Y at *t* if and only if ______? (Wasserman 2004b, p. 694) In metaphysics, there is still some dispute about how to adequately fill out this schema. However, it is widely agreed that a necessary condition for X (or the Xs) to constitute Y is that the relation of constitution that holds between X (or the Xs) and Y involves two *coincidence* conditions. First, material constitution requires *spatial coincidence*: X (or the Xs) constitutes Y at *t* only if X (or the Xs) and Y have the same spatial location at *t*. Second, material constitution requires *material coincidence*: X (or the Xs) constitutes Y at *t* only if X (or the Xs) and Y share all the same material parts at *t* (Wasserman 2004b, p. 694; Wilson 2007, p. 5).

In addition to these coincidence conditions, we can also say something about the formal properties of the constitution relation. Material constitution is often taken as *transitive* and *irreflexive*. Moreover, material constitution is usually considered to be *asymmetric*. Defenders of constitution want to say that *Piece* constitutes *David*, but not *vice versa*. Adding the formal properties of irreflexivity and asymmetry entail a specific view of the constitution relation, that "*constitution is not mere*

---

[19] I discuss the constitution relation in detail in chapter 8, where I scrutinize the causal-constitutive fallacy leveled against the justifiability of EC.

coincidence," (Wasserman 2004b, p. 694; italics in original) for coincidence is both symmetric (i.e., ∀x (Pxx)) and reflexive (i.e., ∀xy (x = y)). As Wasserman says: "This is a substantial commitment, but it is also welcomed by most constitution theorists." (2004b, p. 694)[20]

In addition to the coincidence conditions as well as formal properties of material constitution, the standard view of constitution is commonly thought to hold between *enduring* relata, which are *wholly present* whenever they exist (Wasserman 2004b, p. 708, fn. 3)[21].

2.3.2. Composition[22]

There are different species of composition to be found in the literature both across metaphysics and the philosophy of science. See figure 4 for an overview of these different versions:

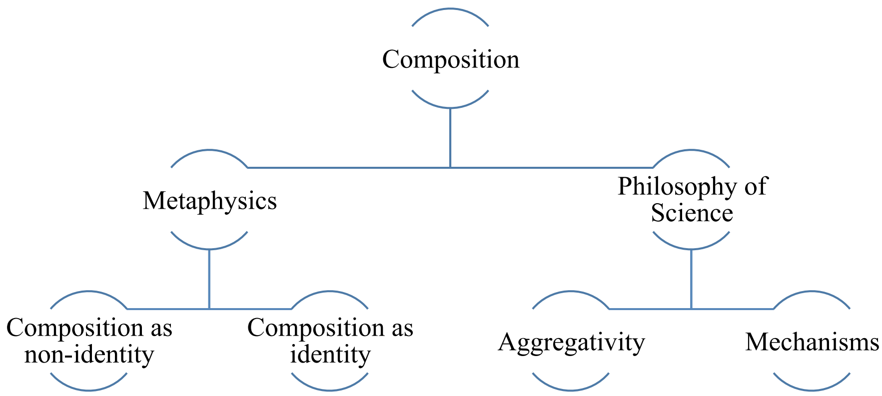

Fig. 4 Overview of different species of composition

[20] For instance, as Lowe states: "Of course, there must evidently be *more* to constitution than just this, not least because constitution is an asymmetrical relation (if *x* is constituted by *y*, then *y* is *not* constituted by *x*), whereas spatiotemporal coincidence is symmetrical." (1989, p. 81; italics in original) Or, as Baker writes: "Pretheoretically, I take the constitution relation to be an asymmetric relation. Piece constitutes *David*; *David* does not constitute *Piece*." (2000, p. 33; italics in original)

[21] I discuss the notion "wholly present" is chapters 3, 4, and 5.

[22] I discuss the composition relation in more detail in chapter 3, 4, and 5.

As it is most commonly understood, "composition is a relation which holds between a plurality of non-overlapping objects, on the one hand, and a single object, on the other […]." (Hawley 2006, p. 2) This general description is consistent across metaphysics and philosophy of science (see e.g., Craver 2007; Craver & Bechtel 2007; Hawley 2010; van Inwagen 1990), although with the exception that not all accounts of composition imply that composition must hold between objects. Composition is also thought to hold between components (Craver 2007; Gillett 2007a). With this small caveat exposed, composition is standardly understood to be a *synchronic* relation of determination. For instance, in his influential *Material Beings*, van Inwagen states explicitly that the verb "compose" in the predicate "the *xs* compose *y*" is to be understood synchronically. As he says:

> "The verb 'compose' in the predicate 'the *xs* compose *y*' is to be understood as being in the present tense, and the same point applies to 'are' in 'are parts of'. Thus, 'are parts of' and 'compose' should be read 'are *now* parts of' and '*now* compose'. Strictly speaking […], our *definiendum* should have been 'the *xs* compose *y* at *t*', and our "primitive" mereological predicate should have been '*x* is a part of *y* at *t*." (1990, p. 29; italics in original)

Or, as Bennett states:

> "Composition is a synchronic or atemporal many-one relation between two distinct objects. It carries with it the cognate notion of 'part': if the *xxs* compose *y*, then each *x* is part of *y* [at *t*]." (2011, p. 81)

Or, as Craver says:

> "At least since Hume, many philosophers have held that causes and effects must be logically independent. If one endorses this restriction on causal relations, then one should balk at positing a causal relationship between constitutively [or compositionally] related properties. Finally, because the [composition] relationship is synchronic, ***Φ'**s* taking on a particular value is not temporally prior to ***Ψ'**s* taking on its value." (2007, p. 153)

In metaphysics, some philosophers argue that the relation of composition is a relation of *identity* (see e.g., Baxter 1988; Lewis 1991; Wallace 2011). Such philosophers think that if X (or the Xs) is (or are) part of Y, then X (or the Xs) is (or are) identical to Y. As van Inwagen expresses this particular view: "*x* overlaps $y =_{df}$ Some one thing is part of both *x* and *y*." (1994, p. 207)[23] That is, if some Z bears identity to both X and Y, it follows that X and Y are identical (van Inwagen 1994, p. 208) One benefit of this view is that if the Xs compose Y, and if the Xs and Y are identical, then this circumvents the problem of having to explain just how parts and whole can be co-located at the same time and in the same place.

For the purpose of this survey, I will not have anything else to say about composition as identity. Indeed, it is not a very popular view in metaphysics (*cf.* Wallace 2011, p. 807) – the reason being that even if the Xs and Y occupy the same region in space-time, and insofar as the Xs compose Y, more often than not there will be some qualitative difference between the Xs and Y (Aizawa & Gillett 2009a). Let us, for example, say that water is composed of H_20. However, whereas water is a solvent, neither hydrogen molecules nor oxygen molecules are solvents. This is a simple case, in that, if H_20 composes water, then the microphysical components and properties are not identical with the macrophysical components and their properties[24]. It is not difficult to construct many more of such examples.

In this sense, and with the exception of composition as identity, the three remaining accounts of composition share with the constitution theorist the view that more than one distinct entity can be co-located in the same place and at the same time. Note, though, that whereas constitution is usually the view that *two* distinct objects can "wholly occupy exactly the same location at the same time […]. [The] similar-sounding 'composition is not identity' usually represents the quite different view that a whole is not identical to its several parts." (Hawley 2010, p. 10)

---

[23] As Lewis puts the point: "A fusion is nothing over and above its parts." (1991, p. 80) Or, as Lewis mentions two pages later: "The 'are' of composition is, so to speak, the plural form of the 'is' of identity." (1991, p. 82)

[24] Opponents of composition as identity include Craver (2007), Hawley (2006) Markosian (1998), McDaniel (2007, 2008), McKay (2006), Merricks (1999, 2005), Sider (2007), van Inwagen (1990, 1994), and Wimsatt (1986), among others.

Furthermore, unlike material constitution, where a relation counts as a metaphysical building relation *only if* it holds between spatially and materially coincident objects, the *nature* of composition (composition as non-identity, aggregativity, mechanistic) is generally considered to consist in what van Inwagen calls:

> "*Composition*: The *xs* compose *y* if and only if no two of the *xs* occupy overlapping regions of space and *y* occupies the sum of the regions of space occupied by the *xs*." (1990, p. 45; italics in original)

Or, as Craver and Bechtel state about composition in mechanisms:

> "Given the compositional relations between mechanisms and their components, the space-time path of the mechanism includes the space-time path of its components. They coexist with one another, and so there is no possibility of their *coming to* spatiotemporally intersect with one another." (2007, p. 552; italics in original)

Composition, in van Inwagen's terms, addresses the "general composition question" (GCQ), in that, it asks about the relationship between Y and the Xs that combine to compose Y. For both van Inwagen and Craver & Bechtel, composition requires *spatial coexistence*: the Xs compose Y at *t* only if Y as a whole shares the same space-time path as the Xs and no two of the Xs occupy an overlapping space-time path. And composition requires *material coexistence*: the Xs compose Y at *t* only if Y as a whole is composed of the Xs and no two of the Xs materially overlap in terms of their parts (Hawley 2006, p. 483). As we saw with material constitution, it is also the case that composition is taken to be transitive, irreflexive, and asymmetric[25].

---

[25] I am aware that Craver & Bechtel (2007) state that composition is symmetric. But it is entirely consistent with Craver & Bechtel's view to argue, first, that composition is asymmetric such as the Xs determine Y, and then, secondly, argue that once instantiated, Y may have top-down effects on the Xs. This is how I understand composition in mechanisms. Furthermore, even though it is standard to understand the relation of composition as being a transitive relation, Bechtel (2009)

To continue our survey of composition, it would be natural to distinguish the GCQ from two other questions one might ask about composition. Van Inwagen calls these for the "special composition question" (SCQ) and the "inverse special composition question" (ISCQ) respectively. The SCQ asks about the circumstances under which the Xs jointly compose Y – as such, the focus is on the constituents, not on the relation between the Xs and Y. The ISCQ, by contrast, focuses on the properties that Y instantiates once Y has been composed by the Xs – here the attention is on the whole rather than either the parts or the part-whole relation[26].

I should stress that it is not my intention to provide the reader with an in-depth analysis of these sorts of composition questions; rather, I use them here for *didactic purposes*. For instance, composition as *aggregativity* informs us both about the SCQ and the ISCQ; similarly for mechanistic composition. However, composition as aggregativity is by far the most *uninteresting* of the two. According to aggregativity, Y is nothing over and above the aggregated sum of the Xs. For example, the mass of a pile of sand is the sum or aggregate property of the masses of the individual sand grains that taken together compose the pile of sand. That is, intersubstitution makes no difference to Y, since the Xs do not interact or overlap in ways relevant to the aggregate Y.

In his work on emergence, Wimsatt has provided an important analysis of aggregate relations. As Wimsatt says: "Four conditions seems separately necessary and jointly sufficient for *aggregativity* or non-emergence." (2000, p. 272; italics in original) These conditions are, he specifies: "For a system property to be an

---

maintains that compositional relations in mechanisms are non-transitive. That is, in a three-level mechanism, say, Bechtel states that mechanistic explanations are always just one way or one level down such that if level 0 explains level 1 and level 1 explains level 2, it does not follow, according to Bechtel, that level 0 explains level 2 (for a similar point, see Aizawa (2013).

[26] In metaphysics, the SCQ has been given by far the most attention (Hawley 2006). For instance, when Markosian asks: "Under what circumstances do some things compose, or add up to, or form, a single object?" (1998, p. 211), he is explicitly addressing the SCQ. The same holds for van Inwagen, who spends most of his time discussing the SCQ in his (1990).

aggregate with respect to a decomposition of the system into parts and their properties, the following four conditions must be met:

> Suppose $\mathbf{P(S_i)} = \mathbf{F\{[p_1, p_2, \ldots, p_n}(s_1)], [\mathbf{p_1, p_2, \ldots, p_n}(s_2)], \ldots, [\mathbf{p_1, p_2, \ldots, p_n}(s_m)]\}$ is a composition function for system property $\mathbf{P(S_i)}$ in terms of parts' properties $\mathbf{p_1, p_2, \ldots, p_n}$, of parts $s_1, s_2, \ldots, s_m$. The composition function is an equation – an inter-level synthetic identity, with the lower level specification [of] a realization or instantiation of the system property.
>
> 1. **IS** (*Inter Substitution*) Invariance of the system property under operations rearranging the parts in the system or interchanging any number of parts with a corresponding numbers of parts from a relevant equivalence class of parts […].
> 2. **QS** (*Size Scaling*) Qualitative similarity of the system property (identity, or if a quantitative property, differing in value) under addition or subtraction of parts […].
> 3. **RA** (*Decomposition or ReAggregation*) Invariance of the system property under operations involving decomposition and reaggregation of parts […].
> 4. **CI** (*Linearity*) There are no Cooperation or Inhibitory interactions among the parts of the system which affect this property." (2000, pp. 275-276)

As I mentioned, examples of aggregative composition are rarely interesting. From the perspective of the ISCQ, Y is the aggregate sum of the Xs, and according to the SCQ there would seem to be nothing of importance about the circumstances in which grains of sand compose a pile of sand. Adding a grain of sand to a pile of sand makes Y one subunit heavier. If we move all the Xs from location Z to location P, it has no qualitative influence on Y. Replacing one X with an equally weighted replica has no effect on the weight of Y, and so on.

According to defenders of mechanistic explanation, scientific explanation in terms of *mechanisms* turns, in part, on the idea that most fundamentally levels of mechanisms are a species of compositional, or part-whole, relations (Craver 2007; Craver & Bechtel 2007; Gillett 2013; Machamer et al. 2000). As Craver specifies:

“Levels of mechanisms are levels of composition. […]. The interlevel relationship [that holds between acting entities at different levels] is as follows: X’s *Φ*-ing is at a lower mechanistic level than S’s *Ψ*-ing if and only if X’s *Φ*-ing is a component in the mechanism for S’s *Ψ*-ing.” (2007, p. 188)

Mechanisms are collections of entities (Xs) and activities (Φs) organized so as to produce regular changes from start-up to termination conditions (Machamer et al. 2000, p. 3), with the organization of mechanisms explaining how the Xs and their Φ-ing are organized to produce something Ψ (Craver 2001, p. 58).

Mechanisms are *hierarchically organized*, integrating different levels together in an explanation of a mechanism. Craver, for example, thinks that the circulatory system is hierarchically organized, in that, the activities Ψ of the circulatory system S are manifested, implemented or instantiated by the heart’s different Xs and their Φ-ing – e.g., the activity of the heart’s pumping blood, the kidney’s filtration of blood, and the venous valves’ regulation of the direction of blood flow (2001, p. 63). That is: “The relationship between lower and higher mechanistic levels is a [compositional] part-whole relationship with the additional restriction that the lower-level parts are components of (and hence organized within) the higher-level mechanism.” (Craver 2001, p. 63) According to this story, the compositional relation that holds between mechanisms and their components implies that the relata do neither spatially nor materially *coincide*, in the sense (i) that the mechanism and its components share the same space-time path, (ii) none of the components occupy the same space-time path, and (iii) none of the components materially overlap[27].

A natural way to contrast aggregativity with mechanistic composition is to look at the part-whole relation in mechanistic composition, since it is often argued by the mechanists that the relation between the Xs and Y is one of *mutual manipulability* (Craver 2007, p. 152) or *mechanistically mediated effects* (Craver & Bechtel 2007, p. 547). Craver provides both an informal gloss and a more formal characterization of mutual manipulability. Formally: “(i) *x* is part of S; (ii) in the conditions relevant to the request for explanation there is some change to Xs *Φ*-ing that changes S’s *Ψ*-ing; and (iii) in the conditions relevant to the request for explanation there is some

[27] Recall that constitution utilizes coincidence conditions, whereas composition turns on coexistence conditions.

change to S's *Ψ*-ing that changes *Φ*-ing." (2007, p. 152) One can change the *explanandum* by intervening to change to Xs, and one can change the *explananda* by intervening to change Y. That is, it is possible to change the *Ψ*-ing of the circulatory system by changing the Xs *Φ*-ing, and *vice versa*. Thus, mechanistic composition yields an account of the three composition questions. In terms of the SCQ, the parts are related to the whole such that one cannot replace, change or subtract one of the parts without affecting a qualitative or quantitative change in S. At the same time, and as an answer to the ISCQ, the whole is related to the parts in such a way that one cannot change the value of S's *Ψ*-ing without changing the value of the Xs and their *Φ*-ing.

2.3.3. Realization[28]

Realization – just like constitution and composition – is an ontological relation: one entity or property (or sets of entities or properties) ontologically depends on one another[29]. The dependence relation in realization is standardly taken to be synchronic or noncausal such that if P realizes Q, P synchronically determines Q (Bennett 2011). As Polger testifies to: "The dependence is non-causal and, to use Amie Thomasson's terms (1999), existential and constant." (2010, p. 195) Or, as Polger & Shapiro put it: "[…] the property instances of wholes are noncausally determined by the property instances of their parts." (2008, p. 219) Furthermore, and again in Polger's words, "[…] ontological dependence relations of this sort are normally asymmetric and irreflexive [as well as] invariant and counterfactual supporting." (2010, p. 195)

The concept of realization entered analytic philosophy in Putnam's (1960/[1975]) classical paper "Minds and Machines," formulating the relationship between the mental and the physical as one of realization based on an analogy to the relationship between the physical arrangement of matter and the abstract operations

[28] I discuss the realization relation in much more detail in chapter 6.

[29] Some accounts of realization discriminate between which kinds of entities are related to realization, whereas other accounts do not (Polger 2010). Although I use "entities" above, the familiar view in philosophy of mind takes realization to hold between states or properties at the level of the mental and the physical, respectively.

of a Turing machine implemented by that physical arrangement of matter. So, Putnam drew the distinction between the logical description of a Turing machine, on the one hand, and the physical states realizing the states referred to by that logical description, on the other, with the idea being that mental states are realized by physical brain states in just this sense (Wilson 2004a, p. 101). Alongside this claim, two further claims were made by Putnam: the first of these was that mental states thus realized can be multiply realized by physical states; and second, the claim that one can (in principle) identify mental states with physical brain states. How the latter of these two claims were rejected on the basis of the first is nicely summarized by Wilson: "Within a few years, the first of these ideas, that of multiple realizability of mental states, had become a central reason for rejecting the second of them, the mind-brain identity thesis […]." (2004a, p. 101) From the rejection of the mind-brain identity thesis, driven by the idea of multiply realizability, arose the view of functionalism about the mind (Block 1980). With the rise of functionalism, the idea that mental states are realized in physical states became the received view on the mind-body relationship (Wilson 2004a, p. 101)

Despite Putnam's groundbreaking contributions to the metaphysics of realization, there is still controversy over the question of whether realization is an ontological dependence relation (Gillett 2002, 2003; Polger 2004, 2007; Polger & Shapiro 2008), whether realization supports ontological reductions (Fodor 1974, 1997; Kim 1989, 1992; Lewis 1972; Shoemaker 2007), whether realization counts against the reduction of higher-level properties to their lower-level realizers (Block 1997; Gillett 2003; Shapiro 2000; Sober 1999), or whether realization can be shown to be inherently context-sensitive (Wilson 2001, 2004a, 2004b). Hence, we should not be surprised to find more than one version of realization in the literature. For simplicity, and nothing else, I shall restrict my survey of realization to the versions listed in figure 5:

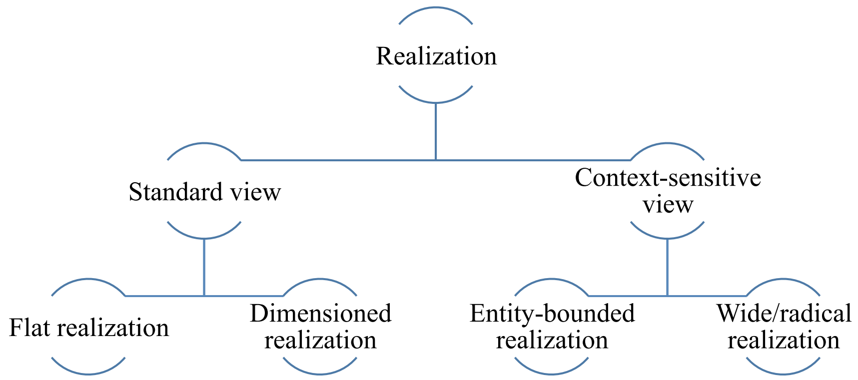

Fig. 5 Overview of different species of realization

It is often said that realization is a relation that holds between first-order properties and second-order properties (i.e., functional properties) such that for P to realize Q is, roughly, for P to play the Q-role (Bennett 2011, p. 82). This is a sketch of what Gillett refers to as *flat* realization (2002). Gillett calls the flat view of realization the standard view, while he thinks of his *dimensioned* account as an alternative, non-standard view of realization. Here I lump both the flat and the dimensioned view under the label "standard view". Why I do so will become clearer, I hope, as I unfold this section.

According to Gillett, the flat view of realization is comprised of two interconnected claims. The first is that the realizer/realized properties are instantiated within *one and the same individual*. In his work on realization, Wilson dubs this claim the *constitutivity thesis*: "realizers of states and properties are exhaustively physically constituted by the intrinsic, physical states of the individual whose states or properties they are." (2001, p. 5) Gillett cites both Kim (1998) and Shoemaker (1999) as proponents of this first claim of the flat view. Consider, e.g., what Kim states here: "It is evident that *a second-order property and its realizers are at the same level … they are properties of the very same object*." (1998, p. 82; italics in original) The second claim of the flat view concerns the causal powers of the realized/realizer properties. Wilson refers to this claim as the *sufficiency thesis*: "realizers are metaphysically sufficient for the properties or states they realize." (2001, p. 4) Consider this time what Shoemaker says: "[…] property X realizes property Y just in case the conditional powers bestowed by Y are a sub-set of the conditional powers bestowed by X […]." (2001, p. 78) Hence, the flat view of realization is based on the ideas that the realized/realizer properties are internal to one and the same individual, and that the causal powers individuative of the realized property match the causal powers contributed to it by the realizer property.

Gillett, however, thinks that the flat view of realization is inadequate; thus his development of what he calls the *dimensioned* view of realization. Let us start with the observation made by Polger & Shapiro that "the dimensioned approach characterizes realization in terms of composition whereas the flat approach views realization in terms of occupiers of functional roles." (2008, p. 213) That is, rather than a one-one relation between property instances, the dimensioned account presupposes a many-one relation between different properties. At the same time, and this is the real culprit, Gillett proposes that there are genuine cases of realization which violate the constitutivity thesis as well as the sufficiency thesis of the flat view. That is, the realized/realizer property may be instantiated within different individuals and the realizer/realized properties may contribute distinct causal powers. In support of the dimensioned view of realization, Gillett considers the example of a diamond, stating that the proponents of the flat view:

> "[…] must deny that the alignment and bonding of particular carbon atoms realizes the hardness of the diamond. For neither (I) [the constitutivity thesis] nor (II) [the sufficiency thesis] is true in this case, since the properties/relations of the carbon atoms are instantiated in different individuals, and contribute distinct causal powers, from the properties of the diamond." (2002, p. 319)

Consider, now, Wilson's *context-sensitive* view of realization, which he taxonomizes into the following three modes of context-sensitive realization: (a) entity-bounded realization; (b) wide realization; and (c), radically wide realization. *Entity-bounded realizations* of P by some lower-level state or process X are total realizations, with both the core and non-core parts being completely located within the individual bearer, B, who has P. According to Wilson, the mammalian circulatory system, which is made up of different parts such as the heart, the arteries, the capillaries, the venules, the blood, and so on, is an example of entity-bounded realizations. Entity-bounded realizations, I submit, are consistent with Gillett's dimensioned view. As Polger & Shapiro point out: "the dimensioned view of realization amounts to no more than a specification of mereological composition. According to Gillett, if objects $O_1 - O_n$ compose an object M, then properties $F_1 - F_n$ of $O_1 - O_n$ realize property H of M." (2008, p. 219) That is, if the heart, arteries, etc., compose the circulatory system, then the properties of the parts realize the circulatory properties H that M has at any given time. However, just as the property "high blood pressure" is not instantiated in any of the individual parts of the circulatory system, similarly for the property of "being hard" in a diamond. Indeed, the hardness of a diamond is a realized property at a higher level than its component properties. Crucially, however, whereas Gillett's dimensioned view of realization is manifested between higher- and lower-levels occupying the same region of space-time (e.g., the space-time region occupied by the diamond), Wilson's notions of *wide* and *radically wide realization* breaks with this restriction. According to Wilson, wide realization $=_{df}$ "a total realization of P whose noncore part is not located entirely within IB, the individual who has P." (2004a, p. 111) An example of wide realization is *fitness* in a biological sense. As Wilson says:

> "[An] organism's *fitness* is its propensity to survive and reproduce in its environment; we can represent the former as a probability between 0 and 1 (the organism's *viability*), and the latter as a number greater than or equal to 0 (the organism's *fertility*) where this number represents the organism's expected number of offspring. In either case, although fitness is a dispositional property of individual organisms […], this disposition is not individualistic, since physically identical organisms may differ in fitness because they have been or are located in different environments: the number that represent viability and fertility may vary *solely* because of an organism's environmental location." (2001, p. 13; italics in original)

In this case, what it metaphysically sufficient for the fitness level of a given organism is not wholly instantiated within that organism – that is, what Wilson calls the total realization of fitness outstrips the boundaries of the individual and includes part of that individual's environment. For the sake of completeness, Wilson defines what he calls *radically wide realization* accordingly: "a wide realization whose core part is not located entirely within [B], the individual who has P." (2004a, p. 116) Examples of radically wide realizations include social actions, involving engagement in cultural practices that themselves have additional social and institutional background conditions, e.g., writing a cheque, voting, withdrawing money from a bank, etc. (Wilson 2001, pp. 13-14; 2004a, p. 116).

2.3.4. Supervenience[30]

Supervenience is primarily "used non-temporally [*viz.*, synchronically], to signify a metaphysical and/or conceptual determination-relation; […], the idea being that something supervenient […] – is "grounded by" – that on which it supervenes." (Horgan 1993, p. 555) Although philosophers disagree about a great many things concerning supervenience, e.g., is supervenience an entailment relation? Is it a relation of explanation, and so on; supervenience is usually understood to be a modal relation such that if a set of A-properties supervene on a set of B-properties there could be no difference in A-properties without there being a difference in B-properties (see e.g., Davidson 1970; Horgan 1982, 1993; Kim 1984, 1988, 1990, 1993, 1998; Lewis 1983, 1986; Stalnaker 1996; etc.). That is to say that a set of A-properties supervene on a set of B-properties if and only if no two things can differ with respect to A-properties without also differing with respect to their B-properties. The slogan "there can't be an A-difference without a B-difference" is applied both to particular individuals and to possible worlds, and across these different qualifications, one may find weak and strong articulations of each view (see figure 6)[31]:

---

[30] I provide a detailed discussion of supervenience in chapter 7.

[31] Philosophers also argue over the modal strength of supervenience. For instance, if the mental supervenes on the physical, which is the received view in the philosophy of mind, there is still substantial debate about whether the supervenience relation between the mental and the physical holds with metaphysical or nomological necessity. For example, Chalmers (1996) has argued that zombies – creatures that are physically indiscernible from human beings, yet who do not have conscious experiences – are metaphysically possible despite most philosophers agreeing that zombies are nomologically impossible (McLaughlin & Bennett 2011)

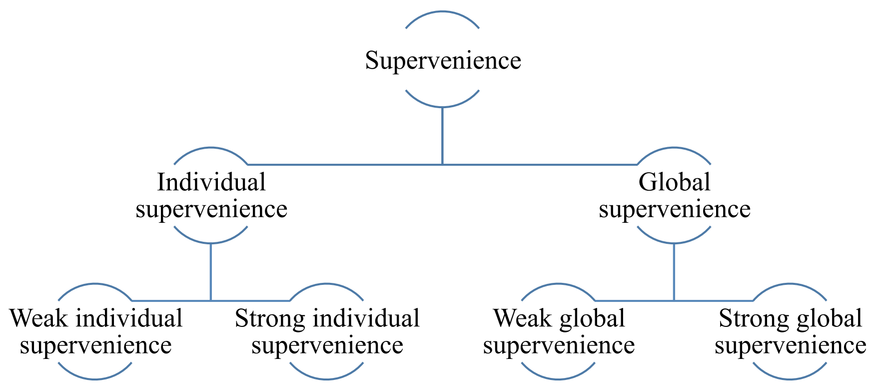

Fig. 6 Overview of different species of supervenience[32]

Let A and B be two sets of properties. Using the modal necessity operator "□", Horgan expresses weak and strong forms of individual supervenience as follows[33]:

[32] I am aware that there are other accounts of supervenience in the literature, e.g., Horgan's (1982) version of *regional supervenience*. I do not consider this account here, because it still expresses a commitment to synchronicity, and will, therefore, not present any problems for my overall aim in this chapter.

[33] These two definitions are formulated by means of modal operators rather than over possible worlds. For a definition of weak and strong individual supervenience defined by means of quantification over possible worlds, consider the following definitions offered by McLaughlin & Bennett: "*A*-properties *weakly supervene* on *B*-properties if and only if for any possible world w and any individuals x and y in w, if x and y are *B*-indiscernible in w, then they are *A*-indiscernible in w." (2011, p. 14; italics in original) And: "*A*-properties *strongly supervene* on *B*-properties if and only if for any possible worlds w_1 and w_2 and any individuals x in w_1 and y in w_2, if x in w_1 is *B*-indiscernible from y in w_2, then x in w_1 is *A*-indiscernible from y in w_2." (2011, p. 14; italics in original) See also Kim (1984, 1987) for a discussion of the difference between strong and weak individual supervenience when either expressed in terms of modal operators or quantified over possible worlds.

"*Weak Supervenience*:

$\Box\ (\forall x)(\ \forall F \in A)\{x \text{ has } F \rightarrow (\exists G \in B)[x \text{ has } G \ \&\ (\forall y)(y \text{ has } G \rightarrow y \text{ has } F)]\}$ (Necessarily, if anything has property *F* in *A*, there exists a property *G* in *B* such that the thing has *G*, and everything that has *G* has *F*." (1993, p. 566; italics in original)

"*Strong Supervenience*:

$\Box\ (\forall x)(\ \forall F \in A)\{x \text{ has } F \rightarrow (\exists G \in B)[x \text{ has } G \ \&\ \Box\ ((\forall y)(y \text{ has } G \rightarrow y \text{ has } F)]\}$ (Necessarily, if anything has property *F* in *A*, there exists a property *G* in *B* such that the thing has *G*, and *necessarily* everything that has *G* has *F*." (1993, p. 567; italics in original)

The difference between these two forms of supervenience relations is that weak supervenience pertains only to entities (or properties) that occupy the *same* possible world, whereas the strong supervenience claim is quantified across *possible worlds*. That is, the strong supervenience claim states that "for any worlds *w* and *w'* and any things *x* and *y* (in *w* and *w'* respectively), if *x* in *w* is *B*-indiscernible from *y* in *w'*, then *x* in *w* is *A*-indiscernible from *y* in *w'*." (Horgan 1993, p. 567; italics in original) In contrast, the weak supervenience claim says that *within* any world, if X and Y are B-indiscernible, and if B supervenes on A, then X and Y are also A-indiscernible.

With the exception of Wilson's conception of wide realization, all the building relations surveyed thus far have traditionally been formulated in such a way that they presuppose that a single individual instantiates both the lower-level properties and the higher-level properties. For instance, "that a single individual instantiates both the subvenient property and the supervenient property." (Horgan 1993, p. 570) One example in which this fails is with the property of *being a bank*. The property of *being a bank* is not exhaustively instantiated by the intrinsic or physical property of the bricks and their organization located in some particular place. The building's having this sociocultural property depends (metaphysically) on the social practice of banking. To accommodate such broader cases of supervenience one solution has been to formulate supervenience in terms of whole or entire possible worlds. Kim coined the conception of *global* supervenience to capture this idea. Global supervenience may be formulated as follows[34]:

> "*A*-properties *globally supervene* on *B*-properties if and only if for any worlds w_1 and w_2, if w_1 and w_2 have exactly the same world-wide pattern of distribution of *B*-properties, then they have exactly the same world-wide pattern of distributed of *A*-properties." (McLaughlin & Bennett 2011, p. 18; italics in original)

2.3.5. Emergence[35]

Emergence is a philosophical term of trade. As Kim puts it: "[It] can pretty much mean whatever you want it to mean, the only condition being that you have better be reasonably clear about what you mean, and that your concept turns out to be something interesting and theoretically useful." (2006, p. 548) Approaches to emergence are often divided into two broad categories, that is, those of diachronic and synchronic emergence, see figure 7:

[34] Here is how McLaughlin & Bennett formulate weak global supervenience and strong global supervenience: "*A*-properties *weakly globally supervene* on *B*-properties iff for any worlds w_1 and w_2, if there is a *B*-preserving isomorphism between w_1 and w_2, then there is an *A*-preserving isomorphism between them." (2011, p. 22; italics in original) And: "*A*-properties *strongly globally supervene* on *B*-properties iff for any worlds w_1 and w_2, every *B*-preserving isomorphism between w_1 and w_2 is an *A*-preserving isomorphism between them." (2011, p. 22; italics in original) See also Stalnaker (1996), and Sider (1999) for ways of making this distinction.

[35] I examine emergence in detail in chapter 7.

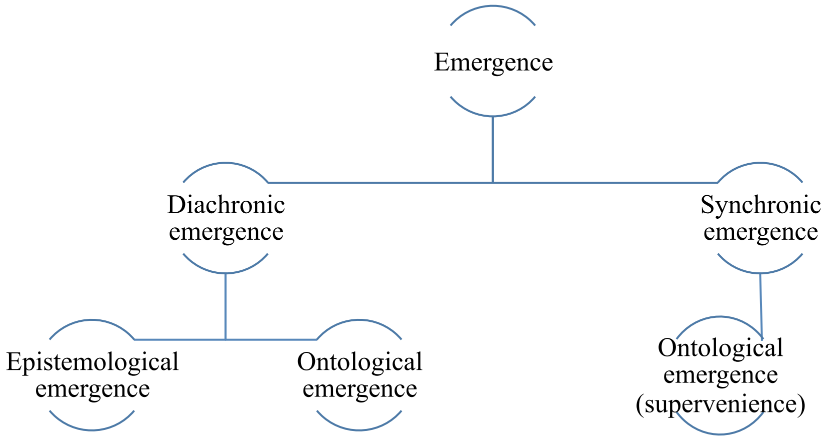

Fig. 7 Overview of different species of emergence

Diachronic emergence is where P emerges as, in part, “a function of some dynamical lower-level or more basic process that unfolds in time. As the system evolves in time new ‘higher-level’ properties will come into being as a function of the unfolding of the more fundamental dynamical process.” (Silberstein 2012, p. 630) Synchronic emergence, by contrast, emphasizes “the co-existence of novel ‘higher-level’ objects or properties with objects or properties existing at some ‘lower-level’.” (Humphreys 2008, p. 431)

Most cases of diachronic emergence are *epistemological* (see e.g., Bedau 1997). Cases of epistemological emergence are cases where emergent properties, say, are merely artifacts of a particular model, and where novel or unexpected properties are commonly unexplainable or unpredictable prior to modeling or analysis. In fact, epistemologically emergent phenomena are phenomena where it is often hopeless to understand the emergent feature of the whole by tracing only the individual parts or processes making up the whole. Often, as Silberstein & McGeever point out, in these cases “we must find a method of representing what the system does on the whole or on average in a manner which abstracts away from causal detail.” (1999, p. 185) Prima facie, at least, dynamical systems theory is a paradigm discipline in which the study of qualitative behavior of higher-level phenomena is dependent on simulation or modeling (Kelso 1995).

Ontological emergence, by contrast, is the idea that higher level phenomena or domains may have causal properties distinct from the causal properties of the constituents, and that those causal properties or capacities may, in the right circumstances, affect the functioning of the lower-level properties or capacities (Mitchell 2012; Silberstein & McGeever 1999). Note, though, that there at least two possible senses of ontological emergence in the literature, and only one of these, I submit, qualifies as a metaphysical building relation.

The first meaning of ontological emergence I call *supervenience emergentism*, which is the standard *synchronic* view of ontological emergence (Broad 1925; Kim 2006)[36]. This is the view that higher-level properties supervene on lower-level subvenient properties, while possessing (the higher-level properties) causal powers that are not entailed by, and consequently not deducible from, the lower-level properties (see e.g., Broad 1925, pp. 67-68). As Horgan states:

> "Certain higher-level properties could be supervenient on lower-level ones […] and also possess the two key features the emergentists stressed: (i) the supervenient higher-order properties could be fundamental causal properties, generating causal forces over and above physical causal forces; and (ii) the connections between lower-order and higher-order properties – supervenience connections – could be metaphysically fundamental, hence unexplainable." (1993, p. 559)

According to Bennett (forthcoming), supervenience emergentism is not a building relation, because of the fact that it implies that emergent properties are *basic* (i.e., fundamental) properties despite the relata involved being connected via the supervenience relation. As Bennett states: […] it follows that ontological emergence […] is not a building relation. Genuinely [fundamental] emergent properties – if there really are any […] – are purportedly no less fundamental than their bases." (Forthcoming, pp. 26-27)

Whether I agree with this or not is unimportant, since this is not the form of emergence that I will defend in the thesis. Fortunately for me, there is a second conception of ontological emergence in the literature, which is less radical than the

[36] Boogerd et al. (2005), Humphreys (2008), and McLaughlin & Bennett (2011) also mention that the standard view of ontological emergence is based on supervenience.

first (or, as some would put it, less strong). This sense of ontological emergence, I call *ontological diachronic emergence*. I explore and defend this notion in chapter 7. The basic idea is that P emerges in virtue of certain dynamical lower-level processes that unfold in time. As these lower-level processes unfold certain higher-level properties emerge. To some this will sound like a rather weak notion of emergence. But, it qualifies as ontological emergence for the following reason: P emerges in virtue of the dynamical activity and organization of the Xs in system S, and the Xs determine the existence of P; once emergent, P may, in the right circumstances, exhibit downward effects on the Xs. Even though P may influence its parts, had the Xs not occurred, P would not had occurred. This second form of ontological emergence is consistent with Silberstein & McGeever's view of ontologically emergent properties: "Emergent properties are properties of a system taken as a whole which exert a causal influence on the parts of the system consistent with, but distinct from, the causal capacities of the parts themselves." (1999, p. 182)

2.4. Conclusion

That concludes my brief survey of five different but familiar metaphysical building relations. I have shown that whereas composition and constitution are usually thought to hold between different objects or components, the relations of supervenience, realization, and emergence are commonly considered to hold between properties of different sorts across levels. Most importantly, I have shown that in their familiar and standard formulations, all of these different building relations are considered as *synchronic* (not diachronic) relations. In chapter 3, I will take a few steps towards formulating DIACHRONIC.

3. Metaphysical building relations: Towards a diachronic account

There is a need to look carefully into just what "time" amounts to when we consider the relationship between metaphysical building relations and temporality. As I said in the introduction, if saying that the diachronic view is important simply because of its emphasis on relations that *unfold over time*, then nearly everyone will accept the diachronic view. The reason for this is that nearly everyone, even proponents of the synchronic view of building relations, give a central place to *causal relations*, and the latter are usually understood to be diachronic (see e.g., Bennett 2011; Craver & Bechtel 2007; Polger 2010; Shapiro 2004). Indeed, if this is all that I intend by the diachronic view, I would be advocating what all of the defenders of the standard view of metaphysical building relations take to be completely familiar, namely that all exponents of the synchronic view include causation into their metaphysics.

What *is* controversial about my diachronic account of building relations, however, is that I give exclusive importance to diachronic relations over and in place of synchronic relations, on the one hand, and that I do not restrict diachronic relations to causal relations, on the other. As I argue in this chapter, and in the chapters to come, metaphysical building relations such as composition, constitution and emergence may all be articulated in accordance with the diachronic perspective. That is the controversial and original claim of the thesis, especially since almost everyone agrees that if there is anything that distinguishes causation, on the one hand, from relations such as composition, constitution, etc., on the other, it is that causation is diachronic, whereas building relations are synchronic.

3.1. Aim

The aim of the chapter is to take a few general steps toward a framework by which to express the diachronic view, before applying and discussing this framework with regards to issues pertinent to EC. Although many of the examples I shall use for this purpose have a distinctly compositional flavor, nothing that I will say in this chapter restricts the diachronic account to the relation of composition. During my discussion, I will keep making use of the standard example from the constitution debate, namely the relation between a piece of marble (*Piece*) and a token statue (*David*) in order to draw out several important points of contrast between the synchronic view and the diachronic view.

In order to avoid complicating things unnecessarily, I postpone any critical discussion of realization, supervenience, and emergence to chapters 6 and 7, and focus just on constitution and composition here.

I present two arguments for the claim that both constitution and composition are, in some cases, diachronic. First, building relations are diachronic in cases where they cannot hold between the very same relata at any particular time instant *t*. Second, building relations hold diachronically in cases where they cannot exhaustively determine the existence of some phenomenon at any particular instant *t*, because neither the parts nor the whole are wholly present at any particular instant *t*.

I am aware that one temporally extended entity might constitute or compose or make up another temporally extended entity, in a way analogous to the standard view of constitution, and as proposed by proponents of the thesis of four-dimensionalism in metaphysics (see e.g., Brogaard 2000). However, failing to press on diachronicity in relation to metaphysical building relations results in theoretical modesty on behalf of theories of building relations. In particular, given that the received view of relations such as constitution and composition is that these relations are synchronic (i.e., durationless) relations, this prevents metaphysical theorizing on composition, and the like, from analyzing time-continuous dynamical systems, which are ubiquitous in nature.

3.2. Overview

In section 3.3, I briefly reiterate the standard view of constitution and composition. In section 3.4, I make a novel distinction between what I call ontological synchronicity and ontological diachronicity in order to clearly separate the synchronic view from the diachronic view, and *vice versa*. In section 3.5, I pursue the first argument for the diachronic view, and in section 3.6, I develop the second argument. Finally, in section 3.7, I add a few additional points of contrast with the synchronic view.

3.3. The standard view of constitution and composition

Michelangelo's *David* is constituted by a piece of marble, *Piece*. This token flag is constituted by a piece of fabric, and so on. These are familiar cases of the constitution view in metaphysics. Although the relation of composition is similar to constitution, the composition relation is usually understood to be a one-many relation between a whole and its parts. For instance, the liquid in this glass is composed of water molecules.

A widespread view amongst constitution theorists (see e.g., Baker 1999; Chappell 1990; Fine 2003; Lowe 1995; Wasserman 2004a, 2009), whatever their other differences, is that the constitution relation between X (or the Xs) and Y must satisfy a number of constraints. If X (or the Xs) constitutes Y, X (or the Xs) and Y exist at the same time instant and share the same material parts (see e.g., Wilson 2007, 2009). In the previous chapter, I referred to these two conditions as the spatial and material coincidence condition, respectively. The standard view of constitution can be easily reformulated in compositional terms by replacing the spatial and material coincidence conditions with the spatial and material coexistence conditions of composition. That is, the Xs compose Y only if Y shares the same space-time path as the Xs and no two of the Xs occupy overlapping regions of space-time, and only if Y is composed of the Xs and no two of the Xs materially overlap in terms of their parts (Craver & Bechtel 2007; Hawley 2006; van Inwagen 1990). Both composition and constitution are usually transitive, asymmetric, and irreflexive. The final condition of the standard view of constitution/composition is *synchronicity*: X (or the Xs) constitutes (or compose) Y only if the relation that holds between X (or the Xs) and Y is synchronic.

The central reason for conceiving of constitution as synchronic is nicely stated by Bennett: "[Metaphysical] building relations do not unfold over time […]. Causation, in contrast, is paradigmatically *diachronic*, and that idea is frequently invoked to distinguish causation from relations like composition, constitution, or supervenience […]." (2011, pp. 93-94; italics in original)[37] That both constitution and composition are understood to hold synchronically is engrained in the very manner in which both relations are articulated. For instance, it is a standard assumption on the part of constitution theorists that constitution requires spatial and material coincidence – X constitutes Y at *t* only if X and Y have the same spatial location at a particular time instant *t* and share the same material parts at that specific time instant *t*. The standard presupposition, then, at the very core of the standard view of constitution, is that the constitution relation holds instantaneously between X (or the Xs) and Y and therefore cannot be a temporally unfolding relation.

Causation, by contrast, may be said to hold between independent events or processes, in the sense that depending on the time interval between the cause and the effect, it is (*prima facie*, at least) possible to think that a cause and its effect never exist simultaneously. As Shapiro says: "[If] C is a constituent of […] P, C exists where and when […] P exists. Thus, for some […] P, if C takes place prior to P's occurrence […], or if C takes place apart from P's occurrence […], then C is not a constituent of P." (2011, p. 160)

[37] Bennett does not endorse this way of distinguishing between constitution and causation. In fact, she immediately rejects this, as she says: "Yet although this is a tempting solution to the problem, occasionally temptation is best resisted. This is one of those occasions. We should not require that building relations be synchronic, because there is at least one important relation that is worth calling a building relation, but that unfolds over time […]." (Bennett 2011, p. 94)

3.4. Ontological synchronicity and ontological diachronicity

In scrutinizing the assumption of the synchronic condition in the standard view of building relations, I start by offering a distinction between *ontological synchronicity* and *ontological diachronicity.* I use the prefix "ontological", here, because I want to stress that when addressing the synchronic-diachronic distinction, we are considering ontological dimensions of this distinction rather than epistemological. If a building relation *R* that holds between X (or the Xs) and Y is ontologically synchronic, it follows that whenever X (or the Xs) is present, Y is present, since for X (or the Xs) to synchronically determine the existence of Y, both X (or the Xs) and Y must be *wholly present* at a particular instant *t*. This is the view that I refer to as the synchronic view. For *R* to be a metaphysical building relation, *R* must hold synchronically between X (or the Xs) and Y. For instance, *Piece* synchronically determines the existence of *David* at time *t* (or at each stage over an interval *t*, …, t_n) if and only if *Piece* and *David* are exhaustively present at *t* or *Piece* and *David* are wholly present at each particular stage over t_1, …, t_n. This synchronic view of building relations is durationless or timeless, in that, the relatum (Y) is determined wholly and completely at a moment in time. In addition to the assumption that *R* is ontologically synchronic such that *R* is wholly present at a particular instant *t*, the standard view of building relations equally presuppose that *R* holds between the *very same token relata* both at a time instant *t* and at each particular stage over an interval, t_1, …, t_n. If we consult the example of *David* and *Piece*, this illustrates that *R* holds between the very same token *Piece* and the very same token *David* at a time or at each specific temporal stage over an interval, t_1, …, t_n.

Contrast this with the conception of *ontological diachronicity*. There are two closely related modes of what I call ontological diachronicity, and both of these modes are implicit in "modern science" cases of building relations, ones far more complex and dynamic than the "commonsense" example of *David* and *Piece* used in analytical metaphysics (see e.g., Ladyman & Ross 2007, p. 21). According to the first sense of ontological diachronicity, if *R* that holds between X (or the Xs) and Y is ontologically diachronic, then *R itself* can never be entirely present at any single moment *t* or at each particular stage over an interval, t_1, …, t_n. On the second sense of ontological diachronicity, if *R* that holds between X (or the Xs) and Y is ontologically diachronic, *R* holds between *relata* that can never be wholly present at a particular time t or at each particular stage over an interval, $t_1, \ldots, t_n$. Ontological diachronicity can be made more evident in exploring some differences between objects (or object-like entities) and processes. As we have seen, objects such as *Piece* and *David* are whole and complete at each moment of their existence – their manifestation is fully determined at a durationless point in time. Processes, by contrast, are *creatures of time* (Noë 2006). Unlike objects, processes are extended over an interval of time – that is, processes are temporally extended in nature. Indeed, the view of ontologically diachronic dependence relations is intended to highlight that if the relata themselves are diachronic – and, thus, temporally extended in nature – then the relation of dependence that holds between such temporally extended relata cannot determine the existence of such relata at a durationless (or atemporal) point in time. In other words, if the relata are inherently temporally extended, can their existence be determined at an atemporal instant? The answer is 'no', as both of the next two argument will establish.

3.5. The first argument for DIACHRONIC

To bring out the first contrast between ontological synchronicity and ontological diachronicity, I begin by asking the following question: In what circumstances, and by which principles, do the microscopic patterns that compose water make up the macroscopic patterns that we typically refer to as water? I begin with the standard assumption that the macroscopic kind "water" is composed of H_2O, that is, of two hydrogen molecules and an oxygen molecule. The interesting question here is whether or not this example is consistent with an ontologically synchronic notion of composition or an ontologically diachronic conception of composition? If H_2O composes water at t, or at each temporal stage over $t, \ldots, t_n$, and if the composition relation is ontologically synchronic, then both of the hydrogen molecules and the oxygen molecule must be wholly present at t or at each moment over $t_1, \ldots, t_n$.

However, according to Ladyman & Ross, this assumption is wrong. That is, water is composed, Ladyman & Ross remind us, "by oxygen and hydrogen in various polymeric forms, such as $(H_2O)_2$, $(H_2O)_3$), and so on, that are *constantly forming*, *dissipating*, and *reforming* over short time periods in such a way as to give rise to the familiar properties of the macroscopic kind water." (2007, p. 21; italics added) Because water is composed in a complex dynamical system it "makes no sense to imagine it having its familiar properties *synchronically*." (Ross & Ladyman 2010, p. 160; italics added) To complicate matters, we cannot even assume that water is composed of H_2O (van Brakel 2010). For instance, in liquid water, water molecules form clusters through interactions, while in ice, as van Brakel mentions, "there aren't really individual molecules." (2010, p. 131) Not even individual hydrogen atoms (H-atoms) are the same over time. In water, van Brakel states, there are both ortho-hydrogen and para-hydrogen H-atoms, and these have quite different physico-chemical properties (2010, p. 132). Fascinatingly, we cannot even assume that water always consists of interactions between H and O atoms. As Belyaev et al. state: "[There] is some probability (however, small) that a water molecule will suddenly transform into a Neon atom." (2001; quoted in van Brakel 2010, p. 132)

What this example indicates is that even if we begin with the assumption that water is composed of H_2O, the dynamics and complexity of the constituent interactions over time contravene the requirement of ontological synchronicity. That is, what is needed is a way of accounting for the dependence relation between water molecules and the macroscopic kind 'water' without presupposing, incoherently, that this relation of dependence is one that holds in an ontologically synchronic manner. Indeed, a synchronic view of composition distorts the complex, inherent temporal dynamics apparent in the microphysical and chemical aspects of water.

3.6. The second argument for DIACHRONIC

The difference between ontological synchronicity and diachronicity also comes about because the former, and not the latter, insists that for X (or the Xs) to synchronically determine the existence of Y, both X (or the Xs) and Y must be *wholly present* at a particular instant. To get a grip on this idea, I consider a couple of additional examples. But let us first see if we can get clear about the notion of "wholly present".

To claim that *David* is wholly present at a particular moment is equivalent to the claim that what there is of *David* at any specific moment is sufficient to determine that *David* exists. Such an account has been proposed by Hofweber & Velleman, though they frame what it is for an entity to be wholly present in terms of identity. As they state: "[An] object *o* is *wholly present* at a time iff the identity of *o* is intrinsic to that time." (2011, p. 55; italics in original) But this way of defining "wholly present" only invites the following question: what is it for an object to have its identity determined intrinsically to a time? Or, consider what Lewis says about the conception of "wholly present": "Let us say something […] *endures* iff it persists by being wholly present at more than one time." (1986, p. 202; italics in original) But this is just as poor a way to characterize what the concept "wholly present" means, in that, it raises another question: what does it mean for an object to be wholly present at more than one time? In the literature, the concept "wholly present" has been (and still is) notoriously hard to provide an adequate definition (Sider 2001)[38]. Here is another try, this time by Wasserman, who states: "x is wholly present at $t =_{df} x$ exists at t and x does not have a proper temporal part at any time other than t." (2004a, p. 77)[39]

These definitions, despite their differences, fit the standard view of building relations. Most defenders of the standard view claim that the sorts of entities under discussion are *enduring* entities, which are wholly present whenever they exist. For instance, if P realizes Q, both P and Q are entirely occurrent, and neither P nor Q is only partially occurrent. Or, if *Piece* constitutes *David*, both *Piece* and *David* are present at the same place at the same time. And so on. But, this is unlike entities such as *events* and *processes* (Hofweber & Velleman 2011; Noë 2006). Consider the following two examples: the process of writing a cheque, and a Mexican wave.

Writing a cheque is a temporally extended process – it takes time from the beginning of writing a cheque to the finished product. Furthermore, insofar as the process of writing a cheque is a temporally extended process, it follows that the process consists of temporal parts each of which involves the laying down of successive drops of ink. As Hofweber & Velleman put the point: "*What there is* of the process at a particular moment – the laying down of a particular drop – is *not sufficient* to determine that a cheque is being written." (2011, p. 50; italics added) Therefore, and in contrast to *David* and *Piece*, for example, "the process [is not

---

[38] Sider's attempt at a definition goes accordingly: "x is *strongly wholly present* throughout interval $T =_{df}$ everything that is at *any* time in T part of x exists and is part of x at *every* time in T." (2001, location 1002/3600; Kindle version)

[39] Wasserman himself thinks that this definition fails as a general definition of "wholly present". But of all the definitions he surveys, this seems to be the one he prefers.

present] in its temporal entirety within the confines of the moment: it is *not fully determined* by the events of the moment to be the process that it is. *Within the moment, it is not all there and it is not fully itself.*" (Hofweber & Velleman 2011, p. 50; italics added) On the identity-based account of "wholly present", it follows that the process of writing a cheque is inconsistent with the requirement that for the Xs to constitute Y, the Xs must be wholly present at a particular time t or at each particular stage over an interval, t_1, …, t_n. Objects, as I have already noted, are durationless, in that, they exist wholly and completely at a time instant. Processes such as writing a cheque, by contrast, are temporally extended. At the beginning of a process or an event, as Noë (2006) reminds us, the process itself has not yet achieved its end state. Similarly, at the end of the process, its beginning is no longer. Indeed, as Noë says, to suppose that "the beginning of an event [and/or a process] would be available, and so present, at its conclusion, […], would be to suppose, confusedly, that events [and/or processes] were in fact object-like structures." (2006, p. 28)

Now, consider a Mexican wave. A Mexican wave is a common occurrence in sports arenas and happens when individual fans stand up slightly after the person next to them does, resulting in what appears to be *a wave running through the crowd*. A Mexican wave might not be a material object – the "material object" being the paradigmatic choice of the composition or constitution theorist with a synchronic bent – but a Mexican wave is *both* extended in time *and* space, and, thus may be conceived as being composed (in some sense) across a time interval. One way to illustrate that the relation R between the Xs (the constituents of a Mexican wave) and Y (the Mexican wave) is diachronically composed is to consider what it implies to claim that R, in this specific example, holds synchronically. Bennett considers the argument that diachronic talk is superfluous, since any diachronic relation of composition can be fully analyzed in terms of synchronic composition[40]. Here is how Bennett schematizes the argument: "The *xxs* at t_1 (or over some interval t_1-t_n) stand in a diachronic composition relation to *y* at t_2 iff the *xxs* exist at t_1, *y* exists at t_2, and at t_2 the *xxs* compose *y*." (Forthcoming, p. 63; italics in original)

The first thing to note about this argument against diachronic composition is that it treats metaphysical relations of determination – such as composition – and

[40] Bennett is, however, quick to reject such an argument.

causal/diachronic relations as two distinct kinds of relations *simpliciter*. That is to say that the composition relation holds between the Xs and Y at t_1 and t_2, and so on, while it is in virtue of certain causal or diachronic relations that the Xs persist through time from t_1 to t_2, and so on. Here *R* is wholly present whenever it holds between the Xs and Y. Thus, if this argument is correct, it entails that the Xs synchronically determine the existence of Y at each moment in time over t_1, …, t_n. The second thing to note about the argument is that it presupposes that *R* holds between *the very same relata* throughout the existence of the Xs and Y. This is easy to understand by looking at the case of *David* and *Piece*. What there is of *David* and *Piece* at each instant is wholly determined at each of those instants. *David* is wholly present at *t* if and only if *David* exists at *t* and does not have a proper temporal part at any time other than *t*.

Let us now apply the critical argument against diachronic composition to the Mexican wave. First, suppose that there are 100 Xs composing Y, and that each X is a token individual with a particular name: Adam, Alice, Betty, John, Michael, Michelle, Richard, Rachel, Will, etc. Second, if the diachronic composition relation can be accounted for in purely synchronic terms, then the composition relation that holds between Y and the Xs at t_2 must also be the composition relation that holds between Y and the Xs at t_1, whereas it is in virtue of some causal or diachronic relation that the Xs persist from t_1 to t_2. If this is correct, it is clear that diachronic relations cannot be compositional (or that compositional relations cannot be diachronic), since diachronic relations would only hold between the Xs, whereas the composition relation holds between the Xs and Y.

However, we can easily reject this skeptical argument. In contrast to the assumption that for *R* to hold synchronically between the Xs and Y, *R* must hold between the very same relata at each particular time instant over the interval it exists, Y continuously *looses and gains* constituents at each moment of its existence and over its career. That is, the composition relation *R*, which holds between Y and the Xs, connects individual parts that do not exist at the same time – in the sense that the Xs are spread out in time and in space – and, as a result, cannot hold between the very same relata at each moment in time over some interval. For instance, Adam, Alice, Betty, John, Michael, Michelle and Richard might stand in a relation to Y at t_1, but not at t_2, since here it is Rachel, Ross, Steven, Stephanie, Will, and Xenia that stand in a relation to Y, and so on until t_n. At the beginning of

Y, Y has not yet achieved its end. Similarly, at the end of Y, its beginning is done with (Noë 2006, p. 28). There should be nothing strange about this, in that, if we grant that Y may gain and loose parts throughout its unfolding, it does not follow that the parts – Adam, say – that are no longer part of the composition relation at t_2, but was at t_1, do not still exist. Of course, it may equally be the case that the parts do not survive the process by which Y comes into existence. Take the process by which wine comes into existence as an example. Here we start with grapes; however, when we get wine, grapes are no longer – strictly speaking – parts of wine. Thus, depending on the case at hand, the Xs may keep on existing or go out of existence.

If there is a composition relation between the Xs (the constituents of a Mexican wave) and Y (a Mexican wave), and because Y continuously looses and gains some of its Xs over time, the relation of composition cannot exhaustively determine the existence of Y at any particular time instant t, since neither the Xs nor Y are *wholly present* at any particular time instant t. If this is correct, we can say the following about the Xs and Y in our example of a Mexican wave: if Y has temporal parts not restricted to any specific time instant t – in that, Y has temporal parts spread out in time and in space – then Y can at best be *partly present* – in contrast to wholly present – at a particular time t or at each particular stage over an interval, $t_1, \ldots, t_n$.

3.7. Status and further reflections

I have presented two arguments for the claim that metaphysical building relations, contrariwise to their standard formulation, may hold diachronically. The first argument was: A building relation holds diachronically in cases where the building relation in question cannot hold between the very same relata. I presented this argument by discussing the composition relation between water and water molecules. The second argument was: A building relation holds diachronically in cases where the building relation cannot exhaustively determine the existence of some higher-level phenomenon at a particular time instant t or at each particular stage over an interval, $t_1, \ldots, t_n$, because neither the parts nor the whole are wholly present at any particular time instant t or at each particular stage over an interval, $t_1, \ldots, t_n$. I presented this argument with the example of a Mexican wave and the process of writing a cheque. I now wish to add a few more points of contrast between the synchronic view and the diachronic view.

3.7.1. Direction of dependence

The standard view of synchronic building relations presupposes that building relations relate entities at different levels in a *vertical*, bottom-up direction of determination, whereas diachronic relations hold *horizontally* between entities at the same level. Bennett provides this diagram illustrating the standard view of synchronic building relations (Forthcoming, p. 45):

$$\begin{array}{ccc} M_1 & \rightarrow & M_2 \\ \uparrow & & \uparrow \\ P_1 & \rightarrow & P_2 \end{array}$$

Fig. 8 Illustration of the assumption that building relations hold vertically, whereas diachronic relations hold horizontally. The vertical axis represents synchronic building (or, relative fundamentality). The horizontal axis represents time.

The case of *David* and *Piece* fits snugly with this diagram. The constitution relation between *Piece* (P_1) and *David* (M_1) holds entirely at t_1 (represented by the vertical arrow). The constitution relation holds between *David* and *Piece* in a *vertical*, bottom-up direction of determination, while is it in virtue of a causal or diachronic relation that *Piece* and *David* persists over time (represented by the horizontal arrow). In other words, the constitution relation between *David* and *Piece* is vertical precisely because it holds synchronically. If this is true, it restricts metaphysical building relations to the vertical axis of the diagram. But, as we have seen, there are cases in which building relations may arguably be said to hold diachronically. On the standard view, however, diachronic relations are not building relations, since diachronic relations are assumed to be horizontal relations, *viz*., that such relations are strictly intralevel relations. If am correct to insist that building relations may be diachronic, it follows that, at least sometimes, the relevant distinction between vertical (synchronic) and horizontal (diachronic) relations is false.

Bennett suggests that we should conceive of diachronic building relations as neither vertical nor horizontal but rather as holding *diagonally*. Consider again the Mexican wave. The Mexican wave, Y, has temporal parts, the Xs, which are not restricted to any specific time instant *t* or to any particular stage over an interval, $t_1, \ldots, t_n$. Indeed, the Xs of Y are spread out over time and in space in such a way that Y and the Xs are precluded from being wholly present at any particular time instant *t* or at any particular stage over an interval, $t_1, \ldots, t_n$. That is, just as Y can only be partially present in time and in space due to having temporal parts spread out in time and in space, the composition relation, *R*, can only be partially present at a particular instant *t*. In fact, we might say, following Bennett, that *R* is vertical; however, whenever *R* is vertical, *R* is only partly present. Similarly, *R* connects temporal parts that are not restricted to one unique time instant *t*, but that unfolds across an interval. In this sense, *R* is operating horizontally – or, across time and space. Because *R* holds both vertically and horizontally, *R*, in the case of the Mexican wave, holds *diagonally*.

3.7.2. Mutually exclusive relations

Insofar as diachronic conceptions of constitution and composition, for example, are justified options, it follows that if conceived of as general frameworks for the study constitution and composition, the synchronically based view and the diachronic view are *mutually exclusive*. That is, it cannot be true *both* that the composition relation *R*, say, that holds between some Xs and some Y is ontologically synchronic *and* that *R* that holds between the very same Xs and Y is ontologically diachronic. Let me qualify this claim. I do not claim that choosing one of the two is equivalent to choosing between two large-scale pictures of the reality of composition. That is, if you opt for one, it does not entail the rejection of the other, *simpliciter*. I am all for plurality. In fact, I have consistently argued that ontological synchronicity, in the case of *David* and *Piece*, for example, is an entirely coherent view. More generally, ontological synchronicity is a live option in cases where the relata are wholly present at some time instant *t* or at each particular stage over an interval, $t_1, \ldots, t_n$. If one is inclined towards pluralism, claiming that ontological synchronicity and ontological diachronicity are mutually exclusive amounts to the following, more limited claim, namely that ontological synchronicity and ontological diachronicity cannot be true of a single phenomenon – *viz*., of one and the same phenomenon in relation to one and the same level of analysis.

3.7.3. Four-dimensionalism

If one were inclined towards four-dimensionalism, one might apply an *atemporal perspective*, within which every constitution or composition relation is synchronic. Four-dimensionalism is the view that everyday objects (including people) are space-time worms that persist through time by having temporal parts none of which are said to be identical with the particular object itself (Brogaard 2000; Sider 2001). However, I am not convinced that adopting the "atemporal perspective" of four-dimensionalism exhaustively supports the general claim that constitution or composition must be ontologically synchronic. The four-dimensionalist might say that in the case of a Mexican wave, Y's temporal parts at *t* are, *atemporally*, part of the larger (so-called) space-time worm that is Y (in its entirety or throughout Y's career). That is, as Sider points out, "we can think of the four dimensionalist's notions of atemporal parthood, and atemporal exemplification generally, as being those we employ when we take an 'atemporal perspective' and contemplate the whole of time." (1997, p. 198) Specifically, according to four-dimensionalism, the statement that the Mexican wave exists is loose talk for the following: "[…] only stages of objects exists, but […] objects have four dimensions in the sense that they have an unfolding temporal dimension in addition to the three spatial ones." (Brogaard 2000, p. 343) Brogaard defines a *stage* as follows: "I take a stage to be an infinitely thin slice of an object along this temporal dimension. No stage is wholly present at more than one time; every stage is wholly present at exactly one time. There is a new stage for every moment at which a given thing exists." (2000, p. 343) Consequently, the four-dimensionalist would say the following about the claim that a Mexican wave is ontologically diachronic: at each moment at which the Mexican wave exists it is wholly present precisely because at each particular point in time, there is an instantaneous stage of the Mexican wave. So the Mexican wave is both wholly present at *t* and over its career (atemporally speaking, of course).

It is not my aim to challenge the four-dimensionalist here – an important task for another time – but it certainly seems to me that insofar as the four-dimensionalist is correct to insist on a synchronic interpretation of a Mexican wave, she may do so, but only because she does so from a *very abstract perspective*, and, as a result, ignores fine-grained temporal details involved in the phenomena under scrutiny. Beyond this quibble, even within the four-dimensionalist perspective, phenomena will vary along many dimensions of differences including homogeneity or heterogeneity of the parts and of the whole, and their respective, overlapping, continuous, and discontinuous "lengths" along the parameter of time. Therefore, the constitution or the composition relation must characterize a very wide variety of phenomena, and the proposed diachronic account helps to sufficiently broaden the metaphysics of building relations to account for more heterogeneous and dynamic phenomena.

3.8. Conclusion

This concludes my development of a few initial steps toward a diachronic view of metaphysical building relations. I now turn to engage in discussions of the diachronic view and the synchronic view – discussions that I think are pertinent to my first thesis aim, namely the metaphysical foundations of EC. In the next chapter, I start by considering the issue of *cognitive assembly* in EC, where I explore a potential relation between the process of cognitive assembly (Clark 2008; Hutchins 2011a) and the special composition question in metaphysics (van Inwagen 1990).

4. The process of cognitive assembly: Further evidence for the diachronic view

When addressing the extended cognition thesis (EC), there are two explanatory targets that one must be careful to distinguish between. The first is the assembly process involved in orchestrating an extended cognitive process and/or an extended cognitive system. I refer to this as *the process of cognitive assembly* (the process of CA). Notice that when addressing the process of CA, we may ask over which timescales, and by which principles, different entities (e.g., certain neural operations, saccadic eye movements, cultural practices, tools) combine to put together or compose a temporally distributed cognitive process and/or system. The other target concerns the newly assembled device once it has been assembled or composed. I call this *the product of cognitive assembly* (the product of CA). In relation to the product of CA, we may ask questions about the difference in properties between the various parts, how information flows and its propagated between the different parts, and how it is that an extended cognitive system instantiates such "processes in ways that ideally solve some problem." (Clark 2008, p. 122)

In this chapter, I attempt to muster additional support for my diachronic account of metaphysical building relations by exploring a recent debate in EC between Clark (2008, 2011) and Hutchins (2011a) concerning the *process of CA*. My independent motivation for this is twofold: first, to sort out some of the positions one might take in this discussion between Clark and Hutchins; and second to defend that at least one of those choices marks a step towards a third-wave version of EC.

4.1. Arguments

When addressing the process of CA, what I call the usual account is as follows: Only processes operating in the here-and-now are responsible for the process of CA and such processes are primarily bodily and neural processes. Prominent advocates of this view include Clark (2008, 2011) and Clark & Chalmers (1998). I should add that the usual account of EC is what has recently been referred to as *first-wave EC*. Defenders of first-wave EC ground their arguments for EC on the ideas that (i) external artifacts are incorporated into the cognitive system of an individual in virtue of the right kind of causal coupling, and (ii) functional similarity between the causal roles of internal and external occupiers. Alternatively to first-wave EC, defenders of *second-wave EC* go beyond parity and focus instead on complementarity between internal and external states and properties (Sutton 2010) and their consequent integration into a cognitive whole (Menary 2007).

In this chapter, I explore one possible route by which to gesture at but also argue for a third-wave version of EC. Specifically, I follow work by second-wave theorists, who are driving particular visions for a third-wave of EC. In particular, I shall propose one approach to the process of CA that exemplifies Sutton's recent gesture towards a third-wave of EC: a version "which dissolves individuals into peculiar loci of coordination and coalescence among multiple structured media [and practices]." (2010, p. 213) This suggestion echoes, I think, what Menary has recently called "enculturated cognition" (EnC). EnC is the "idea that our cognitive abilities are transformed by a cognitive species of cultural practices […]. What we are able to do is augmented and transformed by the acquisition of cognitive practices." (2012, p. 148) Both of these suggestions for a third-wave of EC emphasize the deconstruction of the individual organism as the locus of the process of CA and allow for cultural practices as playing a central role in close coordination with neural and bodily processes.

I aim to unpack this articulation of a third-wave of EC in a way that has not been done before. That is, I will discuss the process of CA in conjunction with work on the relation of composition in metaphysics. I should add that even though the debate about the process of CA is not (strictly speaking) about the metaphysics of composition, those involved in the discussion over CA ask structurally similar questions to those involved in debates about composition in metaphysics. In debates over composition, what is known as the *special composition question* (SCQ) (van Inwagen 1990) concerns the circumstances under which entities assemble or compose another entity. Indeed, both the process of CA and SCQ ask questions concerning the conditions under which entities combine to compose or assemble another entity (or whole). In so doing, both SCQ and the process of CA take as their target the Xs – the constituents – that compose or assemble Y, and analyze the conditions under which the Xs come together to compose or assemble Y.

The first argument that I will develop turns on the fact that composition is understood as a synchronic relation of dependence, and that a synchronic notion of composition is inconsistent with the temporal dynamics inherent in the process of CA. To make this claim, I aim to establish that the restriction of the verb "compose" in the expression "the Xs compose Y" to the *present tense* is metaphysically problematic when considering the nature of time continuous processes such as those involved in the process of CA. This picture, familiar as it is, of X (or the Xs) composing Y at an instant *t*, finds no corresponding image in contemporary debates about the process of CA. In fact, when Clark states that his own targets in *Supersizing the Mind* (2008) are processes operating in the *here-and-now*, nowhere does Clark's temporal qualification "here-and-now" adhere to the assumption that the verb "compose" must be understood to imply "compose at an instant *t*". Consequently, serious inquiry of the process of CA and the SCQ must begin by scrutinizing the actual meaning of the term "now" as it is used to express the claim "the Xs compose Y *now* or in the *here-and-now*". What I shall argue is that the process of CA *must* be stated without implicating a notion of composing or assembling that is synchronic, where "synchronic" means that the Xs are composing Y at an instant *t*. This is important, since I will show that only by problematizing the notion of synchronic composition is it possible to provide a properly motivated answer to the process of CA. That is where the metaphysical action lies in this chapter.

The second argument considers the debate between Clark and Hutchins on the process of CA, with the aim of establishing that as soon as we leave room for the non-trivial role of cultural practices in the process of CA – even when the processes unfold in the here-and-now – this requires that we look beyond the system made up of the individual agent and artifact. That is, we must include into our explanation of the process of CA such features as cognitive norms (Menary 2007) and patterned (cultural) practices (Hutchins 2011a; Menary 2007; Roepstorff et al. 2010). By the end of the chapter, I hope to have established that developing a diachronic account of the process of CA lends support for a third-wave of EC.

4.2. Overview

The structure of the chapter is as follows. In section 4.3, I consider difficulties with giving a satisfactory answer to the qualification that the verb "compose" is to be understood in the present tense – i.e., as *right now* some Xs are composing some Y. In section 4.4, I make use of the outcome of my discussion in section 4.3 to discuss one of Clark's examples of the process of CA: Gray & Fu's (2004) studies on the soft-assembly of interactive microstrategies employed by the brain to solve a given problem. In section 4.5, I consider the debate between Hutchins and Clark on the process of CA. In the final section, I tease out several implications of the discussion in this chapter.

4.3. Discussing the terms "now", "right now" and "here-and-now" [41]

Under what circumstances do a collection of entities compose some further entity? As we saw in the introduction of this chapter, this is van Inwagen's SCQ (1990). Some answer never (see e.g., Rosen & Dorr 2002), some say sometimes, yet only sometimes (see e.g., Markosian 1998; van Inwagen 1990), whereas some say always (see e.g., Lewis 1986; Sider 2001)[42]. In this section, I do not consider any of these options for when (or if) composition holds. Instead, I start by considering difficulties with providing an answer to the assumption that the verb "compose" in the expression 'the Xs jointly combine to compose Y' is to be understood as

---

[41] I use the terms "now", "right now", and "here-and-now" interchangeably.

[42] These different answers express three different attitudes toward composition: first, nihilism, which is the view that there are no conditions under which the Xs compose some Y; second, restrictivism, which states that there are some conditions under which the Xs compose some Y and there are some conditions under which the Xs do not compose some Y; and thirdly, universalism, which is the view that there are no conditions under which the Xs fail to compose something.

meaning ‘*right now*, the Xs combine to compose Y’. Consider, yet again, what van Inwagen says about *tense* and *composition* in *Material Beings*:

> “The verb ‘compose’ in the predicate ‘the *xs* compose *y*’ is to be understood as being in the present tense, and the same point applies to ‘are’ in ‘are parts of’. Thus, ‘are parts of’ and ‘compose’ should be read ‘are *now* parts of’ and ‘*now* compose’. Strictly speaking […], our *definiendum* should have been ‘the *xs* compose *y* at *t*’, and our “primitive” mereological predicate should have been ‘*x* is a part of *y* at *t*.” (1990, p. 29)

On this view, for the Xs to compose Y *now* is to claim that the Xs compose Y at *t*. By my lights, this is a rather elusive way to characterize the idea of “now” with respect to composition, since it leaves it an open question whether van Inwagen means either (i) that composition is a relation between the Xs and Y at time *t*, where the extension of *t* may include ‘over a short interval’ or (ii) that composition is a relation between the Xs and Y at a momentary (i.e., durationless) instant. Consider, however, how Bennett conceptualizes composition in her survey of metaphysical relations: “*Composition* is a synchronic or atemporal many-one relation […].” (2011, p. 81) Furthermore, what we might call the usual account of composition depicts composition as a *vertical relation of dependence* between the Xs and Y, where ‘vertical determination’ is understood to exclude ‘horizontal determination’ (Bennett, forthcoming; Gillett 2007a; Kim 2005). That is to say that if there is a relation of vertical determination between some Xs and some Y, then that relation of determination does not unfold across a time interval. By contrast, if a relation is a relation of horizontal determination such as causation, then the determination in operation is a diachronic (or temporal) kind of determination. So, the usual account of composition – both in metaphysics and in the metaphysics of science – turns on the presupposition that composition is a relation that holds between some Xs and some Y at a durationless instant *t*.

Prima facie, at least, it seems to me that ordinary folk are quite familiar with the idea of “now” or “right now”. For instance, when we talk about something *presently* taking place such as executing a tennis serve, engaging in a conversation with a friend, stirring the pasta sauce, and so on, we (implicitly) appeal to the temporal fact that something – which we may or may not engage in – is happening or taking place *right now*. If the folk are right, which they may or may not be, it would make the set of events to which the term "now" applies include events the temporal duration of which unfolds over milliseconds, seconds, and even minutes. It may only take me a few seconds to stir the pasta sauce, whereas it may take me a few minutes to execute the proper move in a chess game[43].

In metaphysics things are different from the opinions of the folk. For example, in the debate over "now", as we have seen, "now" is usually conceived as "at an instant *t*." A different way of understanding the meaning of "now" is the result of a specific discussion in metaphysics. Following Markosian (2004), it is possible to set up a distinction between two different senses of the notion "X exists now". The first sense is what Markosian calls the *temporal location* sense, where the expression "X exists now" is meant to be synonymous with "X is present". This is the received view of *presentism* in metaphysics (Markosian 2004). The second sense of "X exists now", Markosian calls the *ontological* sense. According to the ontological sense, the expression "X exists now" is understood as shorthand for the claim that X "is now in the domain of our most unrestricted quantifiers, whether it is present, like you and me, or non-present, like Socrates." (Markosian 2004, p. 48) This is the view commonly expressed as *non-presentism*. Presentism and non-presentism make competing claims about temporal ontology (Sider 2006). Presentism is the view that, necessarily, it is always true that only present entities exist, whereas non-presentism is often formulated in an atemporal language that is hostile to presentism (Sider 2001). That is, on the non-presentist view, past and future entities, such as dinosaurs and me 10 years from today, all exist. However, for the presentist, but not the non-presentist, there is something ontologically special about the *now*, in the sense that *only* entities that are currently present, exist.

For my purposes, it matters little whether presentism or non-presentism is ultimately true, in that, all I wish to highlight is that both of these (hotly debated) doctrines in metaphysics are also elusive when it comes to pinning down the precise meaning of the notion "now". First, the presentist states that only present entities exist. But what might this claim amount to in the context of this thesis, which is embedded in the philosophy of cognitive science (broadly speaking)? Consider, for

[43] I should point out that the duration of "now" is also discussed in the philosophy of time consciousness under the heading of *specious present* (see e.g., Varela 1999).

instance, the Hodgkin & Huxley (1952) model of the action potential discussed in Craver (2007).

Action potentials consist of both rapid and fleeting changes in what is known as the electrical potential difference in a neuron's membrane. This electrical potential difference (measured as the voltage difference across the membrane) is known as the membrane potential. The membrane potential consists of a separation of charged ions on either side of the membrane. As Craver specifies: "In the neuron's resting state, positive ions line up against the extracellular surface. In typical cells, this arrangement establishes a polarized resting potential (V_{rest}) of –60 mV to –70 mV. In an action potential, the membrane becomes fleetingly permeable to sodium (Na^+) and potassium (K^+). This allows the ions to diffuse rapidly across the cell membrane." (2007, p. 50)

Unconventional as this example may be in the context of presentism, the question I wish to consider goes to the heart of the ambiguity of how to understand "present" or "now". That is, is the action potential *present* in the presentist sense? Let us start with the (arguably) uncontroversial assumption that insofar as an action potential is present now (whatever we take "now" and "present" to imply), the action potential requires for it to *be* present that it *unfolds over a region of space-time*. In other words, an action potential's nature is such that it is temporally extended in contrast to being completely or wholly present at a momentary instant. This follows from the brute fact that an action potential irreducibly consists of (i), (ii), and (iii) – that is, of (i) a quick increase in mV to a maximum of +35 mV, (ii) a rapid decrease in mV to certain values below V_{rest}, and (iii) a prolonged after-potential period during which the neuron is less excitable – and the manifestation of any of these three stages or parts *take time* (measured in milliseconds). The presentist is committed to the ontological claim that only present entities exist. However, consider some event – the rapid rise in mV to a maximum value of approximately +35 mV – that is happening right now. Too late! That event is over, in the sense that the mV is already rapidly declining to a value below V_{rest}. If we take seriously that it is only present entities that exist, it follows that the first stage of the action potential is now entirely in the *past*. However, according to the presentist, everything that is either in the past or in the future (or both) does not strictly speaking exist; only entities that exist *now* are present.

I do not intend this to be a refutation of presentism; however, if this is indeed one possible outcome of presentism, then it raises a counter-intuitive situation: that the first stage that the action potential consists of is *not* – or, no longer – part of the action potential, because of the fact that the mV is *presently* on the decline. Furthermore, what is *now* part of the action potential – for example, stage two – will, in a very short period of time, cease to be a part of the action potential because it will be entirely in the past. For the cognitive scientist as well as philosophers of cognitive science this result, I suspect, will be unbelievable. For example, in mechanistic philosophy of cognitive science (Bechtel 2008; Craver 2007), lower-level parts compose higher-level entities by the parts being organized in a certain temporal, spatial, and causal fashion. But, the parts themselves will not cease to be part of some higher-level components due to the fact that they operate over different temporal frequencies.

Non-presentism will do no better for my purposes, because non-presentism is stated *tenselessly* or *timelessly*, thus completely ignoring one of the central principles of research across cognitive neuroscience (e.g., Engel et al. 2001; Friston 2010), cognitive psychology (e.g., Ballard et al. 1997; Spivey 2007), as well as dynamical and enactive approaches to cognition (e.g., Clark 1997; Gibson 1979; van Gelder 1998; Varela et al. 1991; Wheeler 2005), namely that cognition *happens in time* (not in some timeless vacuum), and that time *constrains* as well as *limits* the production of cognitive activity. Wheeler captures this emphasis on temporality nicely, as he states:

> "In the psychological arena, such phenomena [i.e., temporally rich phenomena] include (i) the rates of change within, the actual temporal duration of, and any rhythmic properties exhibited by, individual cognitive processes, and (ii) the ways in which those rates of change, temporal durations, and rhythms are synchronized both with the corresponding temporal phenomena exhibited by other cognitive processes, and the temporal processes taking place in the cognizer's body and her environment." (2005, p. 106)

The central problem with both the standard synchronic notion of composition, on the one hand, and the doctrines of presentism and non-presentism, on the other, is that none make concessions to the fact that time is *continuous* (Spivey 2007): one that impedes the treatment of time in terms of arbitrary, discrete step time (t1, t2, etc.)

As I mentioned in the previous chapter, the standard composition theorist is a *synchronic* composition theorist, in that, she/he accepts that the primitive 'X exists at time (or temporal interval) *t*' implies ontological synchronicity with respect to *t*. That is, the synchronic composition theorist claims that if X exists at *t* and if X is part of Y, then X is part of Y at *t* – this will be so no matter how continuous or discontinuous, transient or durable the interval and sub-intervals may be.

This practice of casting the temporal conditions under which X (or the Xs) composes Y into some lockstep or stepwise progression (t1, t2, etc.), involving a sequence of discrete states – such that X_1 composes Y_1 at t_1, X_2 composes Y_2 at t_2, and so on, until t_n – highlights an important difference between temporally complex forms of composition and the kind of composition most metaphysicians have in mind.

To further highlight this difference, consider the following example of returning a tennis serve, in van Gelder & Port:

> "The ball is approaching; you are perceiving its approach, are aware of the other player's movements, are considering the best strategy for the return, and are shifting into position to play the stroke. *All this is happening at the same time*. As you move into place, your perspective on the approaching ball is changing and hence so is activity on your retina and in your visual system … the path of the approaching ball affects which strategy would be best and how you move. *Everything is simultaneously affecting everything else*." (1995, p. 23; italics in original)

In this example, claiming that a system S instantiates Y (returning a tennis serve), and Y is composed of some particular Xs, at a particular point in time *t*, boils down to saying that *during that period of time* Y was composed by the Xs. Keep in mind that we have independent reasons for being suspicious about the term "temporal instant *t*". Consequently, in the process of CA, we should go on to define the relationship between parts and whole as follows: over some period of time, the Xs compose Y, and over that period of time, none of the Xs completely overlap Y.

With respect to the notion that either Y or X (or both) exists *right now*, the standard construal of composition, which implies that "now" or "right now" is to be understood in terms of ontological synchronicity, cannot account for the dynamical and temporally extended phenomena such as returning a tennis serve and an action potential. That is, nothing in the standard account of composition allows for *continuous processes unfolding in real time*. Unlike material objects, which might be timeless in the sense that they exist whole and complete at an instant *t*, processes such as the Mexican wave (chapter 3), an action potential (this chapter), and transactive remembering (the next chapter) are temporally extended in nature. At the beginning of a process, the process as such has not yet achieved its end. Likewise, at the end of a process, its beginning is over. It would obscure the basic temporal nature of processes if processes were supposed, confusedly, to be composed exhaustively at a durationless instant *t*. Importantly, processes such as an action potential, returning a tennis serve, the Mexican wave, etc., are *continuous processes*. Clark provides the following definition of a continuous process:

> "A continuous process is one in which the time-series of explanatorily relevant sub-states cannot be reduced to a sequence of discrete states with jumps in between, but instead requires a genuine continuum of states." (1998b, p. 356)

In the tennis example, it makes little sense to insist that Y is composed *wholly within and only within* each particular sub-interval of *t* and that each transition from one sub-interval to the next involves a complete transition of X and Y such that both X and Y are *wholly present* within one and only one particular sub-interval *at a* time[44].

Similarly in the debate about CA, I submit. For example, when Clark adopts the short-term timescales of the here-and-now to explore just how the brain participates in what (from the perspective of EC) are new distributed cognitive products, Clark

[44] As Spivey mentions: "Real time does not function like a digital computer's clock. It does not move forward and then stop to be counted, and then move forward again only to stop again. At the level of human behavior, real time does not have an objective functional unit." (2007, p. 30)

wants to analyze which neural and bodily processes assemble temporally distributed wholes *right now*. As Clark mentions: "In depicting the processes of on-the-spot recruitment and exploitation as neurally-centered, I meant only to stress the pivotal role, on all these shorter time-scales, of the specifically neural changes that immersion in those cultural practices presumably inculcate." (2011, p. 459) For instance, in *Supersizing the Mind* (2008), most of the case studies referred to by Clark are studies that emphasize the short-term, but varied, temporal scales of bodily and neural operations – timescales that unfold over courses of 50 to 300 milliseconds.

In contrast to the standard view of composition in metaphysics, when Clark states that it is the processes, which operate in the here-and-now that assemble (or, compose) distributed ensembles, what Clark is actually saying is that it is the short-term timescales over which most neural and bodily processes operate that *during that short period of time* assemble or put together some distributed cognitive whole.

Here it is enlightening to consider that the use of "right now" in CA is closely related to how the folk consider the notion "right now". In EC, occurrent distributed cognitive wholes are considered to temporally unfold everywhere from 50 milliseconds and up to a few hours in the case of occurrent emotions and extended instances of decision-making. Similarly, as I argued above, if the folk are right, then the term "right now" would refer to occurrences ranging from 2-3 hours in their entirety to 200-300 milliseconds (and faster). That is, from entire cricket matches to the completion of one saccadic eye movement. Furthermore, as with most (if not all) processes, cognitive processes have subprocesses as well as subphases. For example, the typical time span of a single episode of voluntary biographical remembering is roughly 10 seconds and this trajectory can have any number of both continuous and discontinuous sub-phases and sub-processes (Sutton, personal comm.).

Thus, just as the processes making up a tennis serve return, the processes and subprocesses that compose a token episode of voluntary biographical remembering are *time continuous* – that is, both cases involve continuous processes, and continuous processes and their relevant subprocesses as well as subphases cannot be reduced to a sequence of discrete states instantiated within a discrete temporal slice or stage of time. Spivey expresses this idea, as he says:

> "[Claiming] that a system was in a particular "state," **X**, at a particular point in time, really boils down to saying that the *average* of the system's states during that *period of time* was **X**. This kind of coarse averaging measurement is often a practical necessity in science, but should not be mistaken as genuine evidence for the system actually resting in a discrete stable state." (2007, p. 30; bold and italics in original)

It is, of course, possible to distinguish conceptually between long-term evolutionary timescales, timescales running over developmental and/or cultural-historical time, and the short-term timescales of hours, seconds, and milliseconds (Clark 2011)[45]. But, regardless of how we conceptually carve up time, time is, I submit, continuous (Clark 1998b; Port & van Gelder 1995; Spivey 2007)[46]. It seems highly unlikely that the components assembled on the spot to complete a tennis serve return function in what van Gelder & Port call *arbitrary step time* (t1, t2, etc.). That each new second or millisecond signals, as Spivey puts it "an instantaneous *and simultaneous* updating of the discrete state of each and every unit in the system." (2007, p. 30; italics in original) This complaint applies to the standard view of synchronic composition, in the sense that the standard view delineates time into

---

[45] But these conceptual distinctions may not apply objectively. As Smart (1963), for instance, argues against the A-theory of time, according to which "past", "present" and "future" are understood to objectively apply to the universe, Smart argues that this way of carving up time is an entirely anthropocentric account of time. That is, distinctions of past, present, and future are distinctions made from a particular (human) point of view (Smart 1963, p. 132; see also Sider 2001; for an overview of the A-theory of time, see e.g., Mellor 1998)

[46] Two interesting questions arise at this point: (a) is time in fact continuous? And (b) if time is continuous, is this a necessary or contingent fact about time? I do not discuss (b), because it would take me too far away from the intended topic of this thesis. With respect to (a), I take my metaphysics of time from research in dynamical systems theory and the applications of dynamical systems theory to cognition. I am aware that some philosophers have argued that there is no such thing as "time" (*cf.* McTaggart 1908). Even though this topic is important and worthwhile engaging with, it is beyond the scope of this thesis.

discrete chunks with jumps in between them. In the standard cases, the Xs compose or assemble Y only if the Xs and Y are wholly present at each particular instant at which they exist.

The problem, in short, is that even though both the SCQ and the CA address the question 'under which circumstances, and by which principles' do certain entities compose or assemble other entities, the SCQ is formulated synchronically and, therefore, leaves out the temporal dynamics of actually occurring instances of composition. What we really need is a temporally qualified version of the SCQ; call it the *Temporal Special Composition Question* (TSCQ). Unlike the SCQ, the TSCQ does not presuppose a temporal restriction on the verb "compose" or on the verb "assemble" such that these would imply 'compose or assemble at this very instant'. Instead, the TSCQ asks the question 'over which timescales do processes operate when they jointly compose (or, assemble) a whole?'

In addition, "wholes" may be temporary and a one-off ensemble, a temporary and repeatable one, or something more permanent (Wilson & Clark 2009). Whether it is one or the other is an empirical question – not a question to be settled by metaphysical analysis.

If I am right that the claim 'the Xs compose or assemble Y right now', at least when considering time continuous processes, is an abstraction and should not be mistaken as evidence for the claim that some Xs assemble Y at an ontologically synchronic instant, then the evaluation of empirical evidence supporting my claim must be sensitive to this fact. Fortunately, the empirical evidence is sensitive to this fact. Here, then, is another difference between the understanding of "time" in the standard account of composition, on the one hand, and the kind of composition we find in time continuous systems, on the other. In the former, time is portrayed as a dimension that is *neutral*, i.e., time exists independently of the events or states, or processes, etc., that occur in time, while time (or temporal unfolding) plays a *fundamental role* in the latter. For example, in their discussion of the role of the body (or, embodiment) in cognition, Ballard and colleagues say the following:

> "When the production of intelligent behavior by the body-brain system is taken into account, the constraints of time and space intervene to limit what is possible." (Ballard et al. 1997, p. 723)

If I am correct, synchronic composition treats time only as a specification of the proposition "the Xs compose Y" as taking place at a time. Thus, the locution "P is *v* at *t*" implies that the expression in place of P refers to a proposition (e.g., the Xs compose Y), the expression in place of *v* refers to a truth value (it is either true or false that P), and the expression in place of *t* refers to a particular time instant such that on the standard view "P is *v* at *t*".

To proceed with this temporality-driven critique of synchronic composition, and to hook it up with the discussion of the process of CA, what we need, in the context of the debate over the process of CA, is positive empirical evidence that time really matters for just how and for which processes are assembled in order to solve a given problem[47]. That is, we shall look at an example that Clark argues shows the "balanced use of a set of potentially highly heterogeneous resources assembled on the spot to solve a given problem." (2008, p. 13)

4.4. The process of cognitive assembly: short-term temporal frequencies

Consider Clark's employment of a series of experiments conducted by Gray & Fu (2004) targeting how patterns of interactive behavior emerge at the level of embodiment and how *soft-constraints* – at the embodiment level – determine which of the possible strategies – for solving a given problem – are most likely to be selected given the task environment.

A few points of clarification first: like other researchers in embodied cognition (see e.g., Ballard et al. 1997), Gray & Fu take the notion "embodiment level" to refer to the timescales over which several neural and bodily operations begin to cohere into certain patterns of activity that compose or assemble the bases of interactive behavior. These operations include what Gray & Fu refer to as

[47] One might object to my claim that dynamical cognitive science is incompatible with tenseless accounts of time, in that, you can account for change in tenseless terms as Russell showed (1906, 1946). Briefly, what it is for an entity *E* to undergo change is for *E* to have a property X at *t* and a property Y at t_1 rather than X at t_1. But, notice, if we want to understand (i) the evolvement of the system from *t* to t_1, and (b) how that particular temporal evolvement gives rise to a property difference in E from *t* to t_1, then a synchronic explanation comes up short.

elementary cognitive, perceptual, and action operations that have a typical time-course of 300 milliseconds (2004, p. 362). Thus, in this case, and with respect to the TSCQ, the timescales over which processes operate when they jointly compose or assemble certain cognitive products to solve a given problem is the short-term timescale of 300 milliseconds it takes various neural processes to combine with each other to form a specific "microstrategy" (Gray & Fu 2004, p. 364) – where "microstrategy" refers to patterns of behavior invoked to accomplish a cognitive task. Such microstrategies, Gray & Fu stress, are *softly constrained*, suggesting that there are many possible routes rather than just one (determined route) by which various neural/bodily processes may combine or come together in order to solve a given cognitive task (2004, p. 361).

In the first set of experiments (Gray & Fu 2004), subjects were presented with the task of having to program an on-screen simulation of a VCR control panel. The idea of the experiment was to manipulate the time-course as well as time-cost involved in accessing the information required to program a VCR in order to assess whether the task environment facilitates or discourages the use of "knowledge in-the-world for knowledge in-the-head." (Gray & Fu 2004, p. 364)

In the experiment, subjects were divided into three groups. In the first group (the Free-Access condition), the information was clearly visible in front of the user so that she freely could access the information via saccadic eye movement. In the second group (the Gray-Box condition), the window was partly visible, although the required information (about channel, start time, etc.) was covered with a gray box. To uncover the information the user had to remove the gray box via a mouse click on the gray box. The final group (the Memory-Test condition), who, unlike the others, had memorized all the information required, had to remove the gray box and type in the information. In order to determine the time-course and time-cost involved in each of these three conditions, Gray & Fu analyzed two components: first, the time needed for *perceptual-motor access* to the information; and second, the time needed for *memory retrieval* (see Fig. 9).

What Gray and colleagues found was that time costs of information retrieval, measured in milliseconds, are what determine the combination of processes (biological memory, motor actions, shifts of attention, etc.) assembled to solve the problem. As Clark puts it:

> "[The] subjects settled on whatever strategy yielded (at that phase of the programming) the least cost (measured by time) information retrieval. In fact, they did this even when the fastest mix of resources sacrificed perfect knowledge in the world for imperfect knowledge in the head. Only when the in-the-world data could be accessed with less effort (measured by time) than the data stored in biological memory was it recruited and were calls to the external store "built into" the dominant strategy." (2008, p. 119)

Estimates (in ms) of perceptual-motor and memory retrieval effort by condition		
Condition	Perceptual-motor access	Memory retrieval
Free-Access	500	500-1,000
Gray-Box	1,000-1,500	500-1,000
Memory-Test	1,000-1,500	100-300

Fig. 9 Overview of the time needed to access the information on each condition. Estimates are in milliseconds (adapted from Gray & Fu 2004, p. 368).

If Gray & Fu are correct, what this example clearly indicates is that the psychological phenomena of using softly constrained patterns of information retrieval cannot be appropriately explained without an appeal to richly temporal processes, insofar as various cognitive, perceptual, and motor elements combine to compose such transient microstrategies. In other words, the timescales over which the Xs operate when they assemble Y fail to accommodate the temporally restricted assumption that the Xs do so at *right now* – at this instant t. Indeed, rendering the verb "compose" in "the Xs compose Y" either tenseless (as the non-presentist would insist on) or in present tense (as the presentist would insist on) makes little sense in this dynamical and time continuous domain.

As Clark mentions, temporal "cost-benefit trade-offs are said to provide a soft constraint […] on the mix of motoric, perceptual, and biomemory-based resources that will, other things being equal, be automatically recruited to perform a given information-processing task on a given occasion." (2008, p. 120) That is, Gray & Fu show that the Free-Access group favor perceptual-motor access over memory retrieval, whereas the Memory-Test group favor memory retrieval strategies given that that route of retrieval is much faster than perceptual-motor access. It would seem, therefore, that the timescales over which various cognitive, perceptual, and motor elements combine to compose transient microstrategies are ineliminably *context-sensitive* and will, consequently, differ accordingly to the constraints of the task environment. That is, it is wrong to require that each case of the verb "compose" in the expression "the Xs compose Y" must be true or false once and for all, *viz*., independently of time and context. Processes preclude instantiation at a particular time instant *t*, in the sense that what it is to *be* a process – that is, what it is to persist as a process – involves, necessarily, unfolding over time (Hofweber & Velleman 2011). Therefore, even on the short-term timescales of neural operations, neural processes are not wholly present at any singular instant *t*. For my purposes, then, the real power of the example discussed by Clark is that it shows that even on the short-term timescales over which neural and bodily processes operate, it is ontological diachronicity *all the way down*.

We have here a consequence for those involved in the debate over CA in EC and for metaphysicians with a synchronic persuasion. First, and to repeat what I said in section 4.3, insofar as Clark states that it is the processes that operate *here-and-now* that orchestrate the assembly of hybrid, distributed cognitive wholes, this claim boils down to saying that it is the short-term timescale of bodily and neural processes that during that period of time orchestrated the assembly process of some distributed cognitive whole, and should not be mistaken as evidence for the claim that processes actually assembling some distributed cognitive whole do so at a time instant *t*. As I understand Clark's position – or, the best way to interpret Clark's insistence on the timescale of the here-and-now – is precisely that it is the processes that *unfold over* short-term, but varied, timescales that assemble or compose distributed cognition. Second, if the metaphysics of composition is to apply to dynamical processes involved in the process of CA, then the synchronic account of composition is problematic. Thus, the process of CA must be stated entirely such that it does not implicate assumptions about composition as ontologically synchronic.

4.5. The process of cognitive assembly – distributed over multiple timescales and multiple resources

So far I have argued that any adequate analysis of the SCQ in the context of CA should accept a diachronic conception of composition. I wish now to examine what is presupposed in the argument for the process of CA provided by Clark, namely that it is only the processes operating here-and-now that are responsible for the assembly of distributed cognitive processes and/or systems in conjunction with the claim that those processes primarily responsible for such assembly are bodily and neural processes.

As Clark explicitly states: "My own targets, in the discussion in [Supersizing the Mind] of cognitive assembly, were the processes operating in the here-and-now." (2011, p. 459) And as Clark specifies the kinds of processes in operation in the here-and-now:

> "It is indeed primarily (though not solely) the biological organism that, courtesy especially of its potent neural apparatus, spins and maintains (or more minimally selects and exploits) the webs of additional structure that then form parts of the machinery that accomplishes its own cognizing." (2008, p. 123)

> "Just as it is the spider body that spins and maintains the web that then […] constitutes part of its own extended phenotype, so it is the biological organism that spins, selects, or maintains the webs of cognitive scaffolding that participate in the extended machinery of its own thought and reason. Individual cognizing, then, is organism centered even if it is not organism bound." (2008, p. 123)

There are two assumptions at work in Clark's project. The first assumption is that only processes operating here-and-now are responsible for the assembly of distributed cognitive processes or systems. The second assumption is that the processes most directly responsible for such assembly are bodily and/or neural processes. With respect to the example above, these two assumptions highlight that it is the short-term temporal frequencies at the embodiment level, made up of perceptual, motor, and cognitive processes, which are primarily responsible for the process of integrating appropriate resources into an extended cognitive whole. But, this combination is not the only coherent and live option in the literature. That is, if we call Clark's first assumption "A", and call the second assumption "B", then these assumptions leave open any of the three following combinations: (i) accept (A) and (B) – this is Clark's position in his (2008); (ii) accept (A) but deny (B); and (iii) deny both (A) and (B). That is:

1. Only processes operating here-and-now are responsible for CA, and such processes are bodily and neural processes.
2. Only processes operating here-and-now are responsible for CA, but this does not prevent non-neural and non-bodily processes from significantly contributing to the process of CA.
3. Neither processes operating here-and-now nor bodily and neural processes are primarily responsible for CA.

All three of these combinations have seen defenders in the contemporary literature, although, I suspect, whether there really are exponents of the third combination depends on interpretations of key selected passages. For example, Clark (2011) has attacked Hutchins' "Hypothesis of Enculturated Cognition" (Hutchins 2011a) for presupposing that "cultural practices are *sufficient* to account for all the crucial work of cognitive assembly." (2011, p. 459; italics in original) And as Clark specifies: "I think Hutchins is failing to attend to important differences concerning the shape and timescale of the processes concerned." (2011, p. 459) By stating that Hutchins is failing to attend to different aspects concerning timescales, Clark means:

> "Hutchins' response might be that we should simply reject the conceptual separation between the processes operating on […] various timescales. […]. That is how I read his key suggestion that "both the constraints of cultural practices and the malleable internal microdemons can be seen as elements of a single adaptive system". But while I agree that these are indeed (also) elements of a single long-term adaptive system, that does nothing to diminish the conceptual separation between the long-term evolution of cultural practices, the medium-term effects of my immersion in such practices, and the

> short-term processes by means of which my brain then participates in what (from an extended mind perspective) are new hybrid cognitive routines that productively criss-cross brain, body, and world." (2011, p. 460)

On this interpretation of Hutchins' position, it appears that Hutchins endorses the third combination, namely that neither processes operating in the here-and-now nor bodily and neural processes are what primarily assemble distributed cognitive wholes. Hutchins himself states that one way to avoid the option of combining (A) and (B) is "to abandon the assumption that the biological brain is the essential element. Doing so, of course, requires that one look elsewhere for the apparently impartial forces that assemble cognitive systems." (2011a, p. 439) As Hutchins proposes: "A good start to understanding this process of recruitment would be to notice the role of cultural practices in the orchestration of soft-assembly of extended cognitive systems." (2011a, p. 440)

A worry about the third combination, however, is that its general formulation allows for the following claim: that certain evolutionary conditions for some present cognitive functioning (e.g., from 100,000 years ago) could be actively orchestrating the process of CA here-and-now. I am not aware of any philosophers that have defended such a suggestion, but it is within the logical space of the formulation of the third option.

However, in other passages, Hutchins' own position is much closer to the second option. Clark mentions this possibility as well, as he states: "For as Hutchins himself says, it is only the 'special super-flexible medium' of the brain that *allows* such shared practices to come to orchestrate human learning and response in the first place." (2011, p. 459; italics in original) Or, as Hutchins states:

> "In this perspective, the brain appears as a special super-flexible medium that can form functional subsystems that establish and maintain dynamic co-ordination among constraints imposed by the world of cultural activity, by the body, and by the brain's own prior organization." (2011a, p. 445)

Thus, depending on how one interprets Hutchins' position, it is possible to place him in either option two or three. Even though I say that all three combinations are coherent, it should be clear that not any one of these is free of difficulty. In addition to Clark's criticism of Hutchins' hypothesis of enculturated cognition, Hutchins has argued that the first combination – the one Clark opts for in much of his latest work (Clark 2008, 2011) – is problematic, since endorsing the view that only bodily and neural processes are responsible for the process of CA excludes from view that much of the "heavy-lifting" – as Hutchins is fond of calling it – in the assembly of distributed cognitive ecologies is performed by cultural practices that unfold over longer timescales than those of the here-and-now[48].

A similar sort of ambiguity is present in Clark's authorship, especially when one compares some of Clark's earlier work such as *Being There* (1997) as well as articles such as "Beyond the Flesh: Lessons from a Mole Cricket" (2005a) and "Word, Niche and Super-Niche: How Language Makes Minds Matter More" (2005b) and his latest book *Supersizing the Mind* (2008). In fact, one may also locate a similar kind of ambiguity just by reading through his (2008). The ambiguity consists in the following: whereas Clark, in his earlier work, is much closer to the combination "accept (A) but deny (B)", he is much more in favor of, as we have seen, the combination "accept (A) and (B)" in his latest work. Consider, for example, the following combination of quotes: first, "[my] own targets, in the discussion in [Supersizing the Mind] of cognitive assembly, were the processes

[48] Another critique of the combination of (A) and (B) comes from Rupert (2009). However, whereas Hutchins' objection to this combination is intended to move the EC paradigm towards a more "enculturated" point of view (a third-wave of EC), Rupert's attack of the (A)-(B) combination is meant to show the following: if Clark adopts that combination, rather than supporting EC, the combination "offered is much more in the spirit of an embedded view (Rupert [2004]): the organism is the seat of cognition and locus of control." (2009, p. 47) Whether Rupert's assessment is successful is a difficult question to answer, and controversy rages as I am writing this. But the success of Rupert's argument is beside the point, in that, if it turns out that Rupert is correct, he would still be forced to accept the diachronic account that I am pushing here. As I argued in the introduction, there is no entailment relation from diachronic building relations (such as composition) to EC. Recall, however, that the consequent is not the case, in the sense that if the metaphysical underpinnings of EC are to be underpinned or firmly grounded, this forces EC to accept DIACHRONIC (or some variety of DIACHRONIC).

operating in the here-and-now." (2011, p. 459) And, as Clark specifies, where those processes are primarily located: "Just as it is the spider body that spins and maintains the web that then […] constitutes part of its own extended phenotype, so it is the biological organism that spins, selects, or maintains the webs of cognitive scaffolding that participate in the extended machinery of its own thought and reason. Individual cognizing, then, is organism centered even if it is not organism bound." (2008, p. 123) Here Clark is explicit about endorsing option one – that is, the acceptance of (A) and (B).

But, consider the following couple of quotes, the first from his (2008) and the second from his (1997): first, "[this] is not to deny, of course, that much of the spinning is done by social groups of organisms spread out over long swaths of history." (2008, p. 243; footnote 18) And second: "[The] brain in its bodily context, interacting with a complex world of physical and social structures. These external structures both constrain and augment problem-solving activities of the basic brain, whose role is largely to support a succession of iterated, local, pattern-completing responses." (1997, p. 191) Thus, whereas the first set of quotes puts Clark squarely in the first option, this latter set of quotes puts him firmly in the second option.

It is certainly true that one way to read these ambiguities in both Clark's and Hutchins' work is that they indicate that not any one of the three options can be defended on metaphysical grounds. That is, it is an empirical question just how often and how much of the assembly is performed through the combination of (A) and (B), (A) and not (B) or not (A) and not (B). It is an empirical issue how much and how often the integration is orchestrated internally and how much and how often the integration is assembled externally. This, I suspect, is the correct way by which to understand the ambiguity present in both Clark and Hutchins concerning the process of CA.

However, this raises a different question, namely might there be reasons for favoring one option or combination over the other even though the two combinations are not mutually exclusive? I think that there are such reasons, especially reasons that lend support to the combination "accept (A) but deny (B)". One worry with the first combination is that it threatens to screen-off the fact that even in the here-and-now, history and culture are *always already* embedded and carried along in the practices and artifacts individuals are engaging with (Menary 2007; Sutton 2008, 2010; see also Haugeland 2002). Indeed, the option of taking the second combination on-board is much more in line with a distinctive third-wave of EC theorizing (Sutton 2010; see also Cash 2013; Kirchhoff 2012). That is, even in the here-and-now, across the short-term timescales of hours, minutes, seconds, and so on, the process of CA is not primarily orchestrated by bodily or neural processes but is also sculptured by socially embedded and culturally transmitted practices (see also Sterelny 2010).

As we have seen, Clark's position is (in certain works) consistent with the second option, thereby bringing Clark into contact with a third-wave of EC research. Because of this, I shall consider a case study discussed by Clark on how expert bartenders, when faced with a multiple drink order in noisy and crowded environments are able to successfully solve the problem[49]. Since Clark articulates this example within the framework of niche construction (NC) (Laland et al. 2000), I start by giving a brief introduction to the central tenets of NC. NC, as defined by Laland et al. (2000), refers to:

> "[The] activities, choices and metabolic processes of organisms, through which they define, choose, modify and partly create their own niches. [...]. For example, to varying degrees, organisms choose their own habitats, mates, and resources and construct important components of their local environments such as nests, holes, burrows, paths, webs, dams, and chemical environments." (2000, pp. 132-33)

Organisms adapt to environmental pressures. But organisms also construct, alter, and modify their own environmental niches. Some make burrows, webs, shelters, and other resources. Earthworms are a good example of what is called *pragmatic engineering*. Earthworms engage in burrowing activities, often resulting in a transformation of the structure and chemistry of the soil in which they live (Laland et al. 2000, p. 134). This burrowing activity is important because earthworms, prior to their presence on land, were originally aquatic organisms (Laland 2004, p. 321). As Laland says: only by "co-opting the soils that they inhabit and the tunnels they build to serve as accessory kidneys that compensate for their poor structural

[49] The actual case study is due to Beach (1988).

adaptation […]," (2004, p. 321) can earthworms tackle the physiological demands of a different water- and salt-balance on land.

Many organisms not only alter and transform their environmental niches pragmatically. Much niche construction is a mode of *epistemic engineering*, in the sense that active niche-constructors modify and alter the informational character of the environment (Sterelny 2010, p. 470). For instance, ants lay scent trails between nest and food source. Humans off-load information "onto" the environment to ease the burdens on "internal" memory processing (Donald 1991). Other organisms, like hawks, simply choose the best spot from which to maximize the view of their hunting territory. In a comprehensive study on the "intelligent use of space," Kirsh argues that rearranging spatial relations between environmental resources transforms the problem solving space by reducing the descriptive complexity of the task environment (1995, 2009). Especially in the human lineage, the ramification of epistemic engineering is the establishment of a cumulatively constructed cognitive-developmental niche (Sterelny 2010; Stotz 2010).

Epistemic (and pragmatic) engineering is not only cognition-enhancing in the heat of some problem-solving scenario, since some modifications to the physical and informational environment are transmitted downstream to the following generation. As Sterelny puts this point, "cumulative downstream epistemic engineering" implies transmission of both socio-cultural structures and ecological and technical know-how or expertise enabling the transmission and acquisition of new knowledge (2010, p. 470). In all these cases of NC, both pragmatic and epistemic varieties, what matters, as Laland et al. (2000) emphasize, is that the activity of NC leads to new *feedback cycles*.

In most cases of NC, those feedback cycles run over evolutionary timescales. However, for Clark's purposes, what really matters is that "this whole process has a direct analogue within lifetime learning." (2005a, p. 256) As Clark states:

> "Here, the feedback cycles alter and transform the processes of individual and cultural learning. For example, both educational practices and human-built structures (and artifacts) are passed on from generation to generation in ways that dramatically alter the fitness landscape for individual lifetime learning." (2005a, p. 256)

The example that Clark considers is how an expert bartender, when faced with a multiple drink order in very noisy and crowded circumstances, is capable of solving the problem successfully. Or, alternatively, just how learning within a pre-structured niche with cultural practices and differently shaped glasses, makes it possible for a novice bartender to perform competently. In lifetime learning, or across developmental timescales, the expert bartenders learn how to line up differently shaped glasses in spatial sequences, which, in turn, correspond to the temporal sequence of drink orders. As Clark states: "The problem of remembering what drink to prepare next is thus transformed, as a result of learning within this *pre-structured niche*, into the problem of perceiving different shapes and associating each shape with a kind of drink." (2005a, p. 256; italics added) In this sense, the cultural practices of knowing what to do, and how to do it, when facing a multiple drink order are shaped by the niche constructing activity of previous individuals, and these practices constrain our epistemic access to the world by orchestrating what to attend to and see when so attending (Hutchins 2008). This resonates deeply with Haugeland's idea that normative practices have a certain kind of "normative gravity" (2002, p. 32). In becoming normalized in the practice of bartending, if we take Haugeland's view, what are normalized are not so much behaviors but rather *dispositions to behave*. But even if normal practices or behaviors might never be exactly alike, they are sufficiently alike to be within the same "orbit". Thus, according to Haugeland: "when an individual's dispositions stray from producing behavior within these orbits (that is, types [of normative practices]), they are 'pulled back in'." (2002, p. 32) Another way of articulating this idea is due to the patterned practice approach by Roepstorff et al. (2010). As they mention: "From the inside of a [cultural] practice, certain models of expectancy come to be established, and the patterns, which over time emerge from these patterns, guide perception as well as action." (2010, p. 1056)

Insofar as both of these views are correct, it is not primarily bodily and neural processes that orchestrate a bartender's ability to get the job done but rather neural processes in coordination with normative, cultural practices that come together to assemble such abilities.

4.6. Implications

The combination of ontological diachronicity with the TSCQ implies that whenever the statement 'the processes involved in the process of CA are those processes operating in the here-and-now' is made, it follows that we must understand this statement as expressing 'the processes involved in the process of CA are those processes operating in the here-and-now *during that period of time*'. This particular view, it seems to me, is the implicit view of several philosophers of cognitive science as well as some cognitive scientists themselves (see e.g., Ballard et al. 1997; Beer 2000; Chemero 2009; Clark 1998b; Kirchhoff 2013a; Spivey 2007; Varela et al. 2001; and others). One important implication this has for any synchronic notion of composition – such as the SCQ – is that not only is it ill fitted to analyze temporally complex phenomena; it can never be made to analyze such temporal phenomena *simpliciter*.

Once we make room for a robust diachronic account of the circumstances under which entities of different kinds assemble or compose another entity, and once room has been made for the pivotal role of cultural practices in this process of assembly or composition, it requires (non-trivially) that we look beyond the system made up of the individual agent and artifact. Notice that there is nothing special about endorsing option 1 above: only processes operating in the here-and-now are responsible for the process of CA and such processes are primarily bodily and neural processes. Indeed, prominent advocates of this view include Clark (2008, 2011) and Clark & Chalmers (1998). As we saw in chapter 1, this particular version of EC is what both Menary (2010) and Sutton (2010) refer to as *first-wave EC*. Defenders of first-wave EC ground their arguments for EC on the ideas that (i) external artifacts are recruited into the cognitive system of an individual due to the right kind of causal coupling, and (ii) functional similarity between the causal roles of internal and external physical occupiers. If the arguments for EC focuses on how artifacts are integrated into an individual's cognitive system, then it is not surprising that first-wave versions of EC usually adopt an account of the process of CA along the lines of option 1. That is, in Hutchins' words, "if culture is reduced to a collection of lifeless artifacts" (2011a, p. 444), then the active dynamic processes involved in the process of CA must be bodily and neural. But, if cultural practices may be conceived of as playing a central role in the process of CA, then it follows that some of the active dynamic processes involved in the process of CA lie beyond the system made up of the individual agent and artifact. If this turns out to be correct, then what we have is an account of the process of CA that grounds Sutton's gesture towards a third-wave version of EC: a version "which dissolves individuals into peculiar loci of coordination and coalescence among multiple structured media [and practices]". (2010, p. 213)

I wish to finish this chapter by considering whether it is possible to apply this metaphysical contribution to the debate over the process of CA to settle any disputes in the literature. I think that this is entirely possible. Consider, for example, how Hutchins attempts to push Clark into a strictly neural-oriented position with regards to the process of CA by exploiting an apparent bias in the phrase "on the spot". As Hutchins says:

> "According to Clark, this exploitation happens "on the spot," but the constraints that determine which resources are exploited and how they are related to one another is not entirely formed "on the spot". The "on the spot" phrase highlights the opportunistic nature of cognitive systems. However, without additional discussion, this wording may also bias the solution toward the biological brain by isolating the activity from the context of cultural historical processes." (2011a, p. 441)

One cause for concern about Hutchins' interpretation of the phrase "on the spot" is that there is nothing about the phrase "on the spot" that conceptually entails a commitment to the view that it is the brain that is the most active element in the assembly of distributed cognitive products. An example will make this more concrete. Consider, again, a passage from Hutchins:

> Cultural practices shape active sensing and ways of seeing the world by highlighting what to attend to and what to see when so attending. Clark mentions the activity of seeing a star. A far more interesting example is seeing a constellation, since a constellation exists only by virtue of someone enacting it via a cultural practice that allocates visual attention in a particular way." (2011a, p. 441)

Whichever processes combine to produce the capacity to see a star constellation do so *on the spot* – on the timescales of seconds or, perhaps, minutes. The question is: over which other timescales would such processes be active? Instead of juxtaposing the short-term timescales and long-term timescales (e.g., historical timescales), Hutchins would be better off arguing that there is no problem with depicting the process of CA as unfolding over the short-term timescales of the *here-and-now*, provided that you leave room for the central roles of cultural practices *in* the processes that unfold here-and-now. Insofar as the meaning of "now" is such that it may, in the right circumstances, include a dynamical interval of time, and insofar as the cultural practices within which the cognitive task is carried out unfold within such a dynamical interval of time, then cultural practices may be part of the processes assembling some cognitive ability. That is the real point that one will be able to make by opting for the combination (A) and not (B) above. Consequently, Hutchins cannot appeal to the notion "on the spot" in order to assert that Clark privileges the brain in the process of CA, since the meaning of "on the spot" is entirely contingent – as I have argued – on the time and place of the utterance.

4.7. Conclusion

What I hope to have shown in this chapter is that when considering the process of CA, that is, when we analyze over which timescales certain processes combine to compose distributed cognitive products, there is no ontologically synchronic instant *t* at which this is possible. I wish to finish this chapter by pointing to the following: metaphysical analysis cannot settle the question over which timescales the processes involved in CA are predominantly active. That is a matter of empirical investigation. However, by scrutinizing the metaphysics of what it means for certain Xs to compose a certain Y here-and-now, it is possible, I think, to turn what might look like a purely metaphysical dispute into a productive recipe for empirical research and to set certain constraints for how such research must be carried out.

5. Transactive remembering – diachronic compositional organization

In the previous chapter, I argued that when we consider the process of CA, that is, when we ask over which timescales processes operate when they combine to compose hybrid, distributed cognitive processes and/or systems, there is no ontologically synchronic instant in which such processes are/or systems are instantiated.

Instead of continuing to explore the *process* by which distributed cognitive systems or processes are put together, my aim in this chapter is to analyze what kind of compositional organization certain distributed cognitive systems or processes have during and once they have been established.

The strategy I use consists in offering a case study of *transactive remembering*. My reasons for choosing this cognitive phenomenon are as follows. First, in recent years, transactive remembering has been put to use in justifying EC, especially as a socially distributed mode of EC (Theiner 2009, 2011, 2013; Theiner & O'Connor 2010; Theiner et al. 2010; see also Barnier et al. 2008; Sutton et al. 2010). Second, especially Theiner and colleagues have argued that when people regularly engage in remembering things together – e.g., as intimate couples, families, or work teams do – such collaborative remembering may result in strongly organization dependent modes of socially distributed forms of transactive memory systems.

I do not dispute using transactive remembering to motivate empirical cases for EC. Indeed, I support the view that transactive remembering is a mode of socially distributed cognition (Kirchhoff 2013a, 2013b; Kirchhoff & Newsome 2012). What I *will* dispute is the attempt of Theiner and colleagues to establish the claim that the particular mode of organization in transactive remembering takes the form of mechanistic composition that contravenes Wimsatt's conditions for aggregativity (see e.g., Wimsatt 1986, 2000). I do not dispute this because compositional organization violates Wimsatt's conditions for aggregativity; rather, I dispute this because of the following: (a) mechanistic composition is ontologically synchronic; and (b) the organization of transactive remembering is not well understood by way of synchronic composition.

5.1. Overview

In section 5.2, I provide a short introduction to transactive memory. In section 5.3, I lay out the theory behind mechanistic organization and its relation to composition. While doing so, I illustrate why mechanistic composition is synchronic. In section 5.4, I consider Theiner and colleagues' argument for why transactive remembering or transactive memory systems are mechanistically composed. Finally, in section 5.5, I argue that processes of transactive remembering and the transactive memory systems in which such processes are instantiated are not well understood by way of synchronic mechanistic composition.

5.2. Transactive remembering – a short introduction

Wegner (1987) introduced the concept "transactive memory systems" (TMSs) in an attempt to explain how individual members in long-tenured groups, intimate couples, etc., rely on each other to obtain, process, and communicate knowledge from different domains. Indeed, and as Harris et al. mention, remembering "often occurs jointly in social groups" (2011, p. 268; see also Barnier et al. 2008). As Harris et al. go on to say: "People in close relationships are likely to be behaviorally, emotionally, and cognitively 'interdependent' […] – that is, in collectives such as couples, families, […], and work teams, remembering is an interactive activity where memories are dynamically and jointly constructed […]." (2011, p. 268)

Often, though not always, such dynamically and collaboratively constructed modes of remembering will both emerge from and result in a division of labor and a specialization of knowledge between couples, friends, work teams, and so on. Lewis puts this nicely, when she says: "According to transactive memory theory, group members divide the cognitive labor for their tasks, with members specializing in different domains. Members rely on one another to be responsible for specific expertise such that collectively they possess all of the information needed for their tasks." (2003, p. 587) Transactive memory theory describes both the processes involved in actual instances of transactive memory and the benefits for memory that may occur when remembering is shared between two or more individuals (Barnier et al. 2008; Harris et al. 2011; Lewis 2003; Theiner & O'Connor 2010; Wegner 1987).

A TMS is a species of cooperative and mnemonic division of labor in learning, remembering, and communicating within dyads, triads or larger social groups. For example, Moreland and colleagues have demonstrated that team members trained on the same task tend to develop differentiated but also highly specialized knowledge and are able to collaboratively recall a greater amount of task-specific information than any one individual alone (Moreland et al. 1996; Moreland & Myaskovsky 2000) As Lewis states: "These findings support Wegner's (1987) contention that distributing responsibility for different knowledge domains increases the amount of relevant information available for team tasks." (2003, p. 587)

What are some of the key characteristics of TMSs? According to Moreland and colleagues (see also Lewis (2003) and Theiner & O'Connor (2010)), three factors are central: credibility, specialization, and coordination. That is, different individuals will often possess "different pieces of lower order information relevant to a particular topic […]." (Harris et al. 2011, p. 268) To combine this differentiated information, the members must coordinate their interactions. And, for those interactions to be successful, the individual members must not only rely on one another but must also trust in the specialized knowledge of each member (Lewis 2003, p. 590).

It is helpful to draw a distinction between transactive memory and TMSs (*cf.* Lewis 2003). First, transactive memory is knowledge about the memory system of another individual. In order to benefit from another individual's knowledge, it must be retrieved. As Lewis says, retrieving "the information stored in another person's memory […] depends on transactions (communication, interpersonal interactions) between individuals." (2003, p. 588) In this sense, transactive memory is "meta-knowledge" (Lewis 2003, p. 588) of what the other person knows in conjunction with the transformation of the knowledge oneself has based on that or those transactions. For example, consider this modified version of the example of Otto put forth by Clark & Chalmers (1998) developed by Tollefsen (2006). In this example, Otto is named Olaf and rather than suffering from Alzheimer's as Otto does, Olaf is a philosopher who often gets lost in his work and has a lot of difficulty remembering his appointments, important phone numbers and addresses of colleagues, friends, places, etc. Instead of constantly relying on writing all these things down in his notebook (as in the case of Otto), Olaf turns out to be married to Inga. As Tollefsen says: "Inga [unlike Olaf] has a sharp mind and because they spend a great deal of time together Inga provides Olaf with all of the information that he needs in order to get through his day." (2006, p. 143) Compare this to one of Harris et al's older male participants, who said: "No, I don't use memory aids… Oh hang on, [wife] carries a diary with her all the time… Oh well, if she's got the diary, we're always together and that's it." (Harris et al., forthcoming) Both are examples of an individual's transactive memory. That is, what Olaf is able to remember when it comes to addresses, phone numbers, etc., is deeply dependent on Inga's specialized knowledge, Olaf's trust in that knowledge, and their continued coordination throughout their daily activities.

Inga is an external memory aid for Olaf. But do Inga and Olaf form a coupled system? Adapting Clark and Chalmers' (1998) criteria for when artifacts may be part of some distributed cognitive system, Tollefsen suggests that Inga meets those conditions. First, Inga is always readily available when Olaf needs her, and Olaf typically relies on Inga to accomplish some cognitive task. Second, Olaf trusts the information he acquires from Inga. Whenever Olaf needs information, it is easily accessible, because Inga is always present. Finally, Olaf has typically endorsed the information he receives from Inga at a previous time. Even though Tollefsen argues that long-term couples like Inga and Olaf meet these criteria, and thus qualify as a *bona fide* distributed cognitive system, it is less clear that the species of interaction between Olaf and Inga meets the conditions for establishing a TMS. That is, a transactive memory *system* is manifested when two or more people "cooperatively store, retrieve, and communicate information." (Lewis 2003, p. 588) As Lewis continues: "Whereas *transactive memory* exists in the mind of an individual, a *transactive memory system* exists between individuals as a function of their individual transactive memories." (Lewis 2003, p. 588; italics added) It is evident that Olaf's individual capacity for remembering is augmented by his ongoing interactions with Inga. But in this particular example, it is at best unproven whether Inga's capacities for remembering are augmented or transformed in similar or different ways as well. In a fascinating study on the effects of collaborative remembering in long-married couples, Harris et al. (2011) suggest that there is empirical support for the claim that some long-term married couples might develop and instantiate the existence of a TMS (Barnier et al. 2008; Sutton et al. 2010).

To explore the effects of transactive remembering in long-term couples, Harris et al. (2011) conducted interviews with twelve couples. This procedure was done at their homes over two occasions, one week apart from one another. On each occasion, participants were asked to learn and recall a list of words, to recall various personally relevant semantic information such as the date of their engagement, the names of some of the wedding guests, and to engage in extensive episodic remembering of significant events in their past. In comparing the individual recall data with the data gathered from the collaborative recall tasks, Harris et al. found that certain couples, when compared with the more general semantic descriptions given in the individual interviews, adopted a transactive style of shared remembering. Here is one of the dialogues, where a couple is successfully coordinating a division of cognitive labor in remembering the beginning of their relationship (Harris et al. 2011, p. 291):

Husband: No, I asked her out that night, but she said she couldn't go.
Wife: No, that's right.
H: So then I started to pester her the next week.
W: You did, you turned up after my [classes].
H: [Cooking classes].
W: On Monday night.
H: That'd be it.
W: And took me for coffee.
H: Yes, the next Monday night.
W: And impressed me.
H: Yes.

As Sutton et al. mention about this case: "Compared with the more general semantic descriptions provided in the individual interviews, the joint description of this event in the collaborative interview was emotionally richer and more detailed at a phenomenological and linguistic level, as the couple co-construct an account of his "pestering", and of her being "impressed". (2010, p. 551) In TMSs, therefore, both individuals cultivate one another as external memory aids and develop a "shared system for encoding, storing, and retrieving information." (Wegner et al. 1991, p. 923)

5.3. Composition and mechanistic organization

With this brief introduction to transactive remembering, and before I develop the argument provided by Theiner and colleagues for why TMSs are mechanistically composed, we need a firmer grip on the theory behind mechanistic organization.

5.3.1. Compositional organization of mechanisms

According to defenders of mechanistic explanation, scientific explanation in terms of *mechanisms* turns, in part, on the idea that most fundamentally, relations between *levels* in mechanisms are a species of compositional, or part-whole, relations (Craver 2007; Craver & Bechtel 2007; Machamer et al. 2000). As Craver says: "Levels of mechanisms are levels of composition. […]. The interlevel relationship [that holds between acting entities at different levels] is as follows: X's *Φ*-ing is at a lower mechanistic level than S's *Ψ*-ing if and only if X's *Φ*-ing is a component in the mechanism for S's *Ψ*-ing." (2007, p. 188)

Following from Wimsatt (1986), mechanists distinguish systems that are mechanistically organized, that is, systems exhibiting a strong form of organization dependence, from mere aggregative systems. As Craver states (2001, pp. 58-59): "Suppose that a property Ψ of the whole S is a function of the properties $\{\Phi_1, \Phi_2, \ldots, \Phi_n\}$ of the parts $\{X_1, X_2, \ldots, X_n\}$. Then a Ψ property of S is an aggregate of the Φ properties of Xs when:

(W1) Ψ is invariant under the *rearrangement and intersubstitution* of Xs;

(W2) Ψ remains *qualitatively similar* (if quantitative, differing only in value) with the addition or subtraction of Xs;

(W3) Ψ remains invariant under the *disaggregation and reaggregation* of Xs; and

(W4) There are no *cooperative or inhibitory* interactions among Xs that are relevant to Ψ."

Mechanistically organized systems *contravene* these conditions for aggregativity. In addition to non-aggregativity, mechanisms are *hierarchically organized*. According to Craver, the circulatory system has a hierarchical mechanistic organization, in that, the activities Ψ of the circulatory system S, are implemented by the heart's different Xs and their Φ-ing – e.g., the activity of the heart's pumping blood, the kidney's filtration of blood, and the venous valves' regulation of the direction of blood flow (2001, p. 63). That is: "The relationship between lower and higher mechanistic levels is a [compositional] part-whole relationship with the additional restriction that the lower-level parts are components of (and hence organized within) the higher-level mechanism." (Craver 2001, p. 63) That the compositional relation holds between mechanisms and their components implies that the relata spatially and materially *coexist*. That is: "Given the compositional relations between mechanisms and their components, the space-time path of the mechanism includes the space-time path of its components. They coexist with one another, and so there is no possibility of their *coming to* spatiotemporally intersect with one another." (Craver & Bechtel 2007, p. 552; italics in original)

5.3.2. Levels in mechanisms

In hierarchically organized mechanisms, to say that some entities are compositionally related to entities at another level is to say that those entities are related *vertically*. That is, the entities stand in a relation to one another such that entities at a higher level are dependent (metaphysically) on entities at a lower level. In mechanistic terms: "[An] item X is at a lower level than an item S if and only if X is a component in the mechanism for some activity Ψ of S. X is a component in a mechanism if and only if it is one of the entities or activities organized such that S Ψs.” (2007, p. 548)

The relationship between levels in mechanisms is vertical, because it is *synchronic* and *noncausal*. In contrast to vertical *interlevel* relations, *intralevel* relations are horizontal. Recall from chapter 3, horizontal relations are typically understood to be diachronic relations such as causation. This way of distinguishing between vertical (synchronic) relations and horizontal (diachronic) relations is explicitly endorsed by Craver, as he states:

> “At least since Hume, many philosophers have held that causes and effects must be logically independent. If one endorses this restriction on causal relations, then one should balk at positing a causal relationship between constitutively [or compositionally] related properties. Finally, because the [composition] relationship is synchronic, ***Φ’s*** taking on a particular value is not temporally prior to ***Ψ’s*** taking on its value.” (2007, p. 153)

Because of this, when it comes to the compositional relation between levels in mechanisms, defenders of mechanistic explanation keep intact the assumption that diachronic relations are strictly intralevel, whereas compositional relations hold between levels – in a vertical and synchronic fashion. That the mechanists intend their vertical perspective on composition to be synchronic can be brought further to the fore by exposing how Craver & Bechtel (2007) distinguish between causation, on the one hand, and composition, on the other. Ever since Hume, they note, most, if not all, theories of causation have presupposed that causes and their effects must be *wholly distinct* and that causes (in principle) precede their effects. To underpin this claim, Craver & Bechtel refer to Lewis, who states:

> “C [cause] and E [effect] must be distinct events – and distinct not only in the sense of nonidentity but also in the sense of nonoverlap and nonimplication. It won’t do to say that my speaking this sentence causes my speaking this sentence or that my speaking the whole of it causes my speaking the first half of it; or that my speaking causes my speaking it loudly, or vice versa.” (2000, p. 78; quoted in Craver & Bechtel 2007, p. 552)

Causes and their effects on this account are independent events insofar as an effect may be followed by a cause and, depending on the time interval between the cause and its effect, it is possible (*prima facie*, at least) to imagine that a cause and its effect never exist simultaneously or instantaneously. In contrast to causation, Craver & Bechtel say this about the compositional relation:

> "If a conserved quantity is possessed by one of the components (say, a certain mass or a charge), that conserved quantity is also possessed by the whole. [That is, if] one of the parts bears a mark, that mark is always already born by the whole (by virtue of being born by its parts). The marks do not need to be transmitted upward or downward to have their 'effects;' *their effects are inherited [compositionally], not causally*." (2007, p. 552; italics added)

5.4. Argument for why TMSs are mechanistically composed

In a series of recent papers, Theiner and colleagues (Theiner 2009; Theiner & O'Connor 2010; Theiner et al. 2010; Theiner 2013) have argued that transactive memories are excellent candidates "for *socially manifested* cognitive processes (i.e., cognitive processes of individuals that can be realized only insofar as those individuals participate in groups of a certain kind)." (Theiner & O'Connor 2010, p. 97; see also Wilson 2004a, 2005) I wholly agree with this. However, I disagree with Theiner and colleagues, when they state that the organization of TMSs is compositional in a mechanistic sense (2010, p. 85).

Specifically, Theiner and colleagues argue (amongst other things) that TMSs display their emergent cognitive properties in virtue of their strong organization dependence. Like the mechanists, Theiner and colleagues follow Wimsatt in understanding this species of organization dependence as a failure of aggregativity. Similarly to Craver (2007), Theiner & O'Connor define this species of dependence as follows: "Let s_1 to s_m stand for the m components of a system S (relative to some decomposition D); p_1 to p_n for the n properties of S's components; and F for the organization or mode of interaction between $p_i(s_j)$, such that a system property P(S) is determined by the composition function: $P(S) = F[p_i(s_j)$ for $i = 1$ to n, and $j = 1$ to m]." (2010, p. 84) That is, if P(S) is to count as merely aggregative, P(S) must satisfy the familiar conditions (for a decomposition D of S) provided by Wimsatt (2000, pp. 275-76; reformulated in Theiner & O'Connor 2010, p. 85):

1. **IS:** P(S) is invariant under the inter-substitution of parts of S, or any other parts taken from a relevantly similar domain.
2. **QS:** P(S) remains qualitatively similar (differing only in value) under the addition or subtraction of parts.
3. **DR:** P(S) is invariant under the decomposition and re-aggregation of parts.
4. **CI**: There are no cooperative or inhibitory interactions among parts.

As Theiner & O'Connor go on to claim: "a group S instantiates a cognitive property P(S) just in case P(S) is emergent relative to a decomposition of S into its members, their behavioral and psychological properties, and their modes of social interaction […]." (2010, p. 85)

Here is an example by Theiner & O'Connor (2010) of a three-man team with an established TMS for assembling a radio, where the emergent properties are a mix of partly declarative and partly procedural knowledge about a complex task that none of the individual team members knows how to perform individually (2010, p. 95). The relevant task concerns assembling a radio. In this example, Theiner & O'Connor ask us to imagine the following: "[…] that member A knows how to insert all the mechanical components into the circuit board, B knows how to handle the electronic components, and C knows how to connect each component to all the others in the proper manner." (2010, p. 95) In this particular case, the socially manifested TMS violates the conditions for aggregativity. First, because of their specialized and differentiated knowledge, **IS** is violated. It is not possible that P(S) remains invariant if one or more of the members are inter-substituted. Second, if member A is removed from the job and not replaced with another member with the same specialized knowledge, this contravenes **QS**. **DR** fails, because decomposing the TMS prevents the individuals from transactively engaging with one another. Finally, condition **CI** fails, because "member's awareness of how expertise is distributed affects their individual likelihood of acquiring, recalling, and communicating memory items pertaining to specific categories of information." (Theiner & O'Connor 2010, p. 96)

A similar argument, I submit, can be given for the claim that TMSs in elderly couples (Harris et al. 2011) are mechanistically organized, although this is no part of Harris et al's own investigation (but see Sutton et al. (2010) for an application of such conditions to TMSs). What we could call the *Wimsatt conditions for aggregativity* are all violated in the case of transactive remembering in the elderly couple referred to in section 5.2. First, because of their differentiated ability to recall certain episodic events, condition **IS** clearly fails. Second, condition **QS** fails because the couple succeeds in remembering the events of their first date precisely because of their interactive, dynamic style of collaboration. Third, condition **DR** fails if we remove either the husband or the wife. Finally, condition **CI** fails because without cooperation there would be no TMS. Therefore, all of the Wimsatt conditions for aggregativity equally fail in this case.

5.5. Diachronic composition over synchronic composition

A failure of aggregativity is indicative of an integrated mode of organization. It is an explanatory virtue of applying the Wimsatt conditions for aggregativity to species of socially distributed cognition that this application allows one to emphasize that the type of integration in long-married couples, say, is strongly interdependent. It is from the fact that TMSs implement this particular strong kind of organization dependence that Theiner and colleagues conclude that TMSs are "good candidates for *socially manifested* cognitive processes […]." (2010, p. 97) This exact strategy seems to work best if the Wimsatt conditions for aggregativity are used for explanatory purposes such that we need not be committed to more than the claim that mechanistic explanation in terms of the Wimsatt conditions for aggregativity is an epistemic endeavor. But, there is an underlying *metaphysical* presupposition that I will argue that we need not agree with.

First, the Wimsatt conditions for aggregativity presuppose the existence of a *composition function* such that the emergent properties (e.g., new detail of information, the quality of information, new understanding of previous events in TMSs) of the system can be mechanistically explained in terms of component parts, their activities, and their particular mode of organization. Second, however, the composition function of a system such as a TMS implies that the relationship between the parts and the whole is non-causal and synchronic. It is the fact that the composition function in mechanistic composition is presupposed to be synchronic that I will argue (in the next section) is problematic. I find this assumption problematic for two reasons. I will briefly mention these reasons here, before discussing each one in more detail.

Insofar as the composition function that exists between the Xs and S's Ψ-ing is synchronic, it follows that the Xs and S's Ψ-ing must be *wholly present* at a particular time *t* or at each particular stage across an interval, $t_1, \ldots, t_n$. What else could it mean? Recall from chapter 3, an entity is wholly present at *t* if and only if the identity of that entity is intrinsic to that time. This is the identity-based version of "wholly present" provided by Hofweber & Velleman (2011). A different, although related, version of the concept "wholly present" is provided by Wasserman, who states: "x is wholly present at $t =_{df} x$ exists at t and x does not have a proper temporal part at any time other than *t*." (2004a, p. 77) The key for our purposes is that if an entity is wholly present that entity is exhaustively determined at a specific instant in time, *t*. In this sense, if an entity is wholly present at *t*, it is completely and exhaustively present at a moment in time, *t*.

If the assumption that the composition relation that exists between the Xs and S's Ψ-ing does not imply that the Xs and S's Ψ-ing are *wholly present* at a particular time *t* or at each particular stage across an interval, $t_1, \ldots, t_n$, then the composition function will fail to be ontologically synchronic. We can safely assume that the prefix "ontological" in the notion "ontologically synchronic" is the intended prefix by Theiner and colleagues here, because the metaphysics of composition aims to carve nature "at its joints" (so to speak) rather than merely expressing an epistemological aspect of mechanistic organization. However, in the case of TMSs, the presupposition that the composition function is ontologically synchronic is inconsistent with the fact that in TMSs, S's Ψ-ing is a temporally extended process composed of different parts – given the condition that temporally extended processes fail to be wholly present. It is, of course, important to note that it does not follow from this that processes – even though processes are temporally extended in nature – cannot exhibit certain properties at particular times within that process. This is, however, not what is being disputed. What I am disputing is that S's Ψ-ing can be wholly present at any particular time instant *t* or at any specific temporal moment across an interval, $t_1, \ldots, t_n$. If this is true, the assumption that TMSs have their emergent properties grounded due to a synchronic composition function is problematic.

Emergent properties such as remembering episodic events in TMSs are not dependent on any synchronic composition function for their existence due to the fact that they are *diachronically emergent properties*. As Silberstein says about diachronically emergent properties: "[…] P only emerges as, at least in part, a function of some dynamical lower-level or more basic process that unfolds in time. As the system evolves in time new 'higher-level' properties will come into being as a function of the unfolding of the more fundamental dynamical processes." (2012, p. 630) Because emergent properties, in TMSs, become available over a certain time-course of collaboration, it makes little sense to insist that TMSs have their emergent properties determined at a particular time instant. This is wholly consistent with transactive memory theory, which predicts that new emergent properties of shared remembering will develop over time. As Harris et al. state: "[The] longer a group has shared the encoding and retrieval of information, the more efficient their shared remembering." (2011, p. 272)

5.5.1. Processes in TMSs are temporally extended processes

In their studies on transactive remembering in long-married couples, Harris et al. (2011) found, among other things, that certain implicit processes of collaborative cross-cuing and co-construction may lead to S's Ψ-ing having such emergent properties as new information (in terms of quantity), richer emotional and vivid remembering (in terms of quality), and sometimes to new understandings of the same shared event (Harris et al. 2011, p. 292).

In each case, S's Ψ-ing is a temporally extended process, in the sense that the particular way that Ψ *persists* depends on Ψ unfolding over time. In philosophy, the terms *endurance* and *perdurance* are commonly understood to denote two different ways in which entities persist (Hofweber & Velleman 2011; Olson 2006; Sider 2001). One might think that it is enough to show that Ψ persists by unfolding through time – or, more technically, Ψ's persistence necessarily involves that it extends through a fourth dimension, namely time (Sider 1997) – to show that Ψ persists through time by *perduring* rather than enduring. But, unfolding through time is neutral between perdurantism and endurantism. Thus, it is not enough to establish that Ψ unfolds through time to ground the claim that Ψ persists through time by perduring. If that were all there was to the debate between perdurantists and endurantists nearly everyone would agree. What *does* distinguish perdurantism from endurantism, and *vice versa*, is *how* Ψ is understood to persist through time.

A caveat: I am fully aware that the endurance-perdurance debate is standardly presupposed to be a debate about how *material objects* persist. I will ignore this restriction, since it will help me establish a problem with the assumption that TMSs have their emergent properties synchronically determined in an instantaneous, upward relation of determination.

Consider *the* central difference-maker between perdurantism, on the one hand, and endurantism, on the other, concerning *how* entities persist. On the perdurantist view, entities persist through time by (a) unfolding over time, and (b) at no single (snapshot) instant in time *t* (if there indeed is such an instant) do entities persist by being wholly present at that time *t*. Processes, I have argued, are good candidates for entities that persist by perduring, in the sense that processes, unlike material objects, *are creatures of time* – as Noë says (2006). If we apply S's Ψ-ing to the perdurantist template, it follows that Ψ persists through time by having compositional parts at times other than *t*. That is, Ψ is temporally extended in nature and as such can never be *whole* in the ontologically synchronic sense of 'whole'. On the endurantist view, by contrast, entities persist through time by (a) unfolding over time, and (b) at each time instant *t* being wholly present. In this sense, enduring entities such as material objects are timeless, that is, they exist wholly and completely at a moment of time. Applying this template to our case of S's Ψ-ing, it follows that Ψ persists through time by being exhaustively or wholly present at a particular instant *t* or at each stage or point over a time interval, $t_1, \ldots, t_n$. That is, insofar as Ψ is wholly present at *t*, then Ψ exists at *t* and does not have a temporal part at any time other than *t*. That is the usual account of endurantism (Olson 2006).

If we combine these two different frameworks for how entities persist with one of our examples of TMSs, we get the result that S's Ψ-ing persists by perduring rather than enduring. Nothing about S's Ψ-ing could ever persist in the ontologically synchronic sense of endurantism. As Goldie, for example, puts the issue concerning emotions such as grief: "grief is a kind of process […], which unfolds over time, and the unfolding over time is explanatorily prior to what is the case at any particular moment." (2011, p. 119) Recall from chapter 3, where I made a similar point by reference to the process of writing a cheque (Hofweber & Velleman 2011) and the Mexican wave. Consider the example of transactive remembering in long-term married couples from section 5.2 due to Harris et al. (2011, p. 291):

Husband: No, I asked her out that night, but she said she couldn't go.
Wife: No, that's right.
H: So then I started to pester her the next week.
W: You did, you turned up after my [classes].
H: [Cooking classes].
W: On Monday night.
H: That'd be it.
W: And took me for coffee.
H: Yes, the next Monday night.
W: And impressed me.
H: Yes.

I depict this temporal process of transactive remembering as follows (see figure 10):

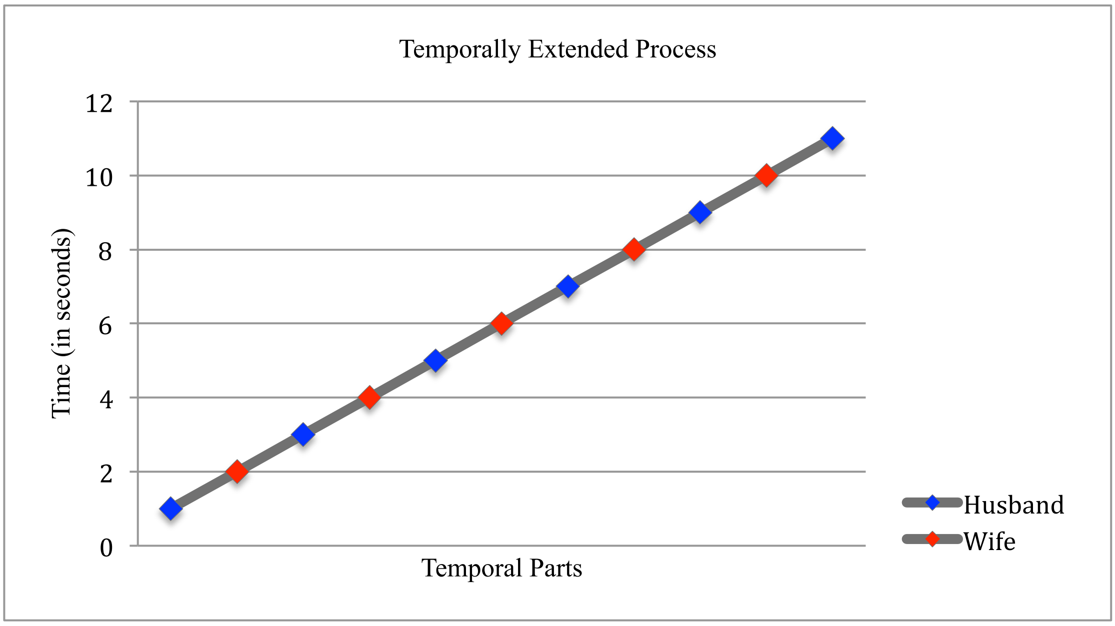

Fig. 10 Diagram of a temporally extended process, where each blue dot refers to the Husband and where each red dot refers to the Wife in the dialogue. For simplicity I represent the entire process to unfold over 11 seconds, with each transaction taking 1 second. Harris et al. (2011) do not depict the process in this way. I use this diagram in order to highlight that a temporally extended process consists of temporal parts.

S's Ψ-ing is a temporally extended process that consists of 11 transactions between husband (blue dot) and wife (red dot) over a time interval. For example, the husband (the first blue dot) states at time 1: "No, I asked her out that night, but she said she couldn't go", while the wife (the first red dot) says at time 2: "No, that's right", and so on until time 11, where the husband (last blue dot) says, "Yes".

When an entity occupies an interval of time by being wholly present in each constituent moment, it is said to endure. Even though some have argued that this conception of endurance is problematic (Hofweber & Velleman 2011), it is the usual account of endurantism. In our example, S's Ψ-ing is a temporally extended process and is decomposable or divisible into parts. The facts that S's Ψ-ing is temporally extended is entailed by its existing throughout a period of time. I have attempted to depict this idea in figure 10 above. But *how* S's Ψ-ing persists is inconsistent with it enduring. To see this, let me begin with an example I used in chapter 3, namely Hofweber & Velleman's (2011) case of writing a cheque.

Writing a cheque is a temporally extended process – it takes time from the beginning of writing a cheque to the finished product. Furthermore, insofar as the process of writing a cheque is a temporally extended process, it follows that the process consists of temporal parts each of which involves the laying down of successive drops of ink. As Hofweber & Velleman put the point: "*What there is* of the process at a particular moment – the laying down of a particular drop – is *not sufficient* to determine that a cheque is being written." (2011, p. 50; italics added) That is, "the process [is not present] in its temporal entirety within the confines of the moment: it is *not fully determined* by the events of the moment to be the process that it is. *Within the moment, it is not all there and it is not fully itself.*" (2011, p. 50; italics added) On the identity-based account of "wholly present", it follows that the process of writing a cheque is inconsistent with the requirement that for this process to persist at a time t or over an interval $t_1, \ldots, t_n$, that process must be wholly present at each particular time instant t at which it exists.

Similarly with TMSs. S's Ψ-ing is a temporally extended process, that is, it unfolds over time (in figure 10, from time 1 until time 11). The fact that it is a temporally extended process equally implies that it consists of temporal parts each of which consists of an individual engaging in an interactive sort of transaction (in figure 10, from the husband at time 1 until the husband at time 11). Following Hofweber & Velleman's example above we can then say: what there is of the process of S's Ψ-ing at a particular instant is insufficient to determine that a collaboratively and transactive species of remembering is unfolding. Or, in other words, what there is of this process at time t is insufficient to determine the existence of transactive remembering insofar as transactive remembering requires for its existence *spatiotemporal continuity*. Thus, and to echo Goldie's statement from above, the unfolding over time of a process is explanatorily prior to what is the case at any one or particular moment. Thus, S's Ψ-ing cannot be wholly present at each particular instant t at which it exists. Consequently, S's Ψ-ing cannot be exhaustively composed at an ontologically synchronic instant t.

I need to say a bit about the notion "temporal part," because as I am construing it here, I am breaking ranks with the usual account of the concept "temporal part" in metaphysics, in the sense that temporal parts are standardly understood to be material objects themselves. This can be illustrated with the fact that the use of temporal parts talk is typically invoked while seeking answers to the traditional paradoxes surrounding the problem of colocated or coincident material objects. Problems that pertain, especially, as we saw in chapter 2, to the material constitution relation.

Here, however, I want to speak of temporal parts with regards to temporal processes, and where the parts are processes themselves. Call the following definition of temporal parthood the usual account (Olson 2006): "x is a temporal part of $y =_{\text{df}} x$ is a part of y, and x exists at some time, and every part of y that does not overlap x exists only at times when x does not exist." (Olson 2006, p. 739; see also Sider 1997, p. 205; 2001, p. 59) In discussing temporally extended processes such as S's Ψ-ing, the usual view of temporal parts implies that insofar as the husband's transaction at time 1 is a temporal part of S's Ψ-ing, say, it follows that the husband's transaction at time 1 *overlaps* everything of S's Ψ-ing at time 1. Whether or not it is fruitful to talk in terms of complete overlap in this case I am doubtful about. However, in order to avoid begging the question against this received view of temporal parts, and in order to take seriously that S's Ψ-ing is a diachronically extended process, I shall give the following definition of what temporal parthood talk amounts to in the case of TMSs. The definition is accordingly: X (the first blue dot in figure 10) is a temporal part of S's Ψ-ing during that period of time (circa one second) $=_{df}$ (i) X unfolds during, but only, during that period of time, (ii) X is part of S's Ψ-ing during that period of time and (iii) X overlaps everything that is part of S's Ψ-ing during that period of time.

5.5.2. Diachronically emergent properties

In the previous section, I deliberately focused on S's Ψ-ing and the temporal parts that make up S's Ψ-ing. In so doing, I engaged in a discussion that was entirely *on the same level* – so to speak – looking only at S's Ψ-ing as a temporally extended process with temporal parts. I did not analyze the interlevel relationship between the emergent properties of S's Ψ-ing and the temporal parts of S's Ψ-ing from which S's emergent properties arise. It is this particular task that I wish to engage in now.

Mechanistic composition – in part underpinned by contravening the Wimsatt conditions for aggregativity – assumes that emergent properties are mechanistically organized in virtue of the mechanisms giving rise to these emergent properties having an ontologically synchronic composition function. But, the presupposition that TMSs have their emergent properties due to the TMS having an ontologically synchronic composition function sits uneasily with emergent properties in TMSs. The reason for this is that emergent properties in TMSs are *diachronically emergent properties*. I consider diachronic emergence in detail in chapter 7. Here, however, I wish to explore the argument that an ontologically synchronic view of composition is incompatible with the diachronic nature of emergent properties in TMSs such as those implemented by long-married couples.

Recall the distinction between transactive memory and TMSs (Lewis 2003). In addition to each individual's transactive memories, TMSs include the kind of processes the individuals bring to bear in order to combine their transactive knowledge (Lewis 2003, p. 590). This is an important point, because not all transactions between two or more individuals result in successful joint remembering. For example, Harris et al. found that certain processes such as the inability to combine and integrate individual memories as well as the use of corrections by one or both of the individuals had negative effects on successful joint remembering (2011, p. 289).

The first implication I want to draw is the following: while the Wimsatt conditions for aggregativity as well as synchronic mechanistic composition emphasize the co-existence of novel 'higher-level' properties with properties and components existing at a 'lower-level', the Wimsatt conditions and synchronic mechanistic composition fail to emphasize the inherently temporal and historical features of TMSs. This follows, I submit, from the fact that one can apply each of the conditions for aggregativity (IS, QS, DR, and CI) to couples failing to successfully engage in joint remembering. In either case of joint remembering, that is, whether or not the enacted processes lead to successful joint remembering, S's Ψ-ing will not remain invariant under intersubstitution (violation of condition IS), will not remain qualitatively similar under subtraction (violation of condition QS), will not remain invariant if decomposed (violation of condition DR), and presupposes, necessarily, cooperation (violation of condition CI). It is therefore not possible to determine whether the compositional organization of TMSs yield successful modes of shared remembering by looking only at such synchronic conditions between higher-level properties of a TMS and its lower-level properties, components, and their organization.

This result is not too surprising, since emergent properties in TMSs come about in virtue of some diachronic and historical patterns of interaction between two or more individuals. Indeed, because emergent properties in TMSs become available over a certain time-course of collaboration, this speaks against the view that TMSs have their emergent properties determined synchronically.

For example, Harris et al. (2011) report that insofar as both the husband and the wife were able to differentiate their expertise, recognize one another as experts at remembering different aspects of episodic events, and given the particular transactive processes enacted, interactive transactions between couples would lead, in the right circumstances, to emergent properties such as remembering *new details* that both individuals could not remember alone as well as *emotionally richer* descriptions of the event(s) (Harris et al. 2011, p. 291). As Harris et al. mention: "[…] this implicit process of collaborative cross-cuing and co-construction seemed to produce elaborations and new information, and to lead to more detailed, episodic, emotionally richer recall." (2011, p. 292) Insofar as emergent properties in TMSs are mechanistically composed, and insofar as collaborative cross-cuing and co-construction are *parts* of the TMS giving rise to emergent properties such as remembering more detailed and emotionally richer descriptions of episodic events, it follows that these properties are emergent due to a *diachronic composition function*. That is, even though the processes giving rise to emergent properties in TMSs operate in the here-and-now, none of these determine the existence of emergent properties in TMSs at an ontologically synchronic instant t, because the lower-level processes are themselves dependent on spatiotemporal continuity for their existence. That is, neither higher-level emergent properties nor lower-level processes giving rise to emergent properties are wholly present at a particular time instant t or at each point over an interval $t_1, \ldots, t_n$. Indeed, the statement that emergent properties in TMSs are emergent in virtue of their compositional organization must not be mistaken for the statement that the lower-level processes are organized in such a way at a particular instant t that they wholly determine the existence of P, some higher-level emergent property.

5.6. Conclusion

I hope to have achieved two things in this chapter. First, I hope to have shown that by providing an analysis of transactive remembering, it is possible to show that insofar as TMSs are concerned, their compositional organization contravenes conditions for synchronic composition. Second, even though my critical target in this chapter has been Theiner and colleagues work on TMSs, I hope that my critical discussion of this work establishes a further development in this area of research.

6. Species of realization and the free energy principle

In the previous three chapters, the focus has been on a diachronic explication of composition either in relation to specific metaphysical presuppositions or in relation to issues and case studies pertinent to EC. In this chapter, and in the chapters that will follow, I start to move beyond the composition relation and to consider relations of realization (this chapter), supervenience and emergence (chapter 7), and constitution (chapter 8).

This chapter considers the realization relation. Especially, I examine the potential confluence between work on the metaphysics of realization in analytical philosophy (see e.g., Aizawa & Gillett 2009a, 2009b; Gillett 2002, 2003, 2007a; Polger 2007, 2010; Shapiro 2004; and Wilson 2001, 2004a, 2004b) and research on the free energy minimization formulation in cognitive neuroscience (see e.g., Clark 2013; Friston 2002, 2003, 2010, 2011; Friston et al. 2012; Hohwy et al. 2008). The free energy principle is a variation of the so-called predictive brain hypothesis, which states that the brain is constantly making predictions about potential future events. It is the growing consensus in theoretical as well as systems neuroscience that a fundamental feature of neural computation is that the brain is always trying to reduce prediction error[50].

In the philosophical literature, no work has yet been done to examine whether these two fields of research can be brought together. In this chapter, I address some

[50] According to the free energy formulation, all biological systems are driven to reduce or minimize an information-theoretic property known as "free energy" (Friston et al. 2012, p. 1). In addition to its information-theoretic rendition, the free energy principle is mathematically similar to physical entropy in thermodynamics (Friston 2011).

of the issues that arise when such a potential confluence is scrutinized. My reasons for exploring this potential confluence are threefold.

First, in his “Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science” (2013), Clark indicates that the free energy principle fits snugly with extended approaches to cognition. That is, active construction of the niche – over the timescales of phylogeny, ontogeny, and the here-and-now – yields extra-neural resources for minimizing prediction error, thus suggesting that perception and action team up with the more slowly evolving backdrop of culturally distributed practices to minimize prediction error. If this is correct, the free energy minimization formulation might well turn out to be a natural ally of EC.

Second, the free energy principle portrays the mind-body relationship as one of free energy minimization, and the most prominent application of the realization relation has been to address the mind-body problem. As a first approximation, this suggests that the realization relation and free energy minimization can be brought together.

Finally, in various writings, Wilson (2001, 2004a, 2004b) has argued that insofar as the relation of realization is metaphysically wide, the realization relation may be used to ground the metaphysics of EC. By extension, then, if EC is consistent with the free energy principle, and if the metaphysics of EC is consistent with wide realization, then the free energy principle is consistent with wide realization. This follows from transitivity.

It is one thing that it follows logically. Whether it follows as a matter of empirical fact is a different issue entirely. Because it is an empirical question whether free energy minimization is *realized* by certain physical entities, and since I aim to provide a philosophical treatment of this question, I shall frame my arguments conditionally. I provide an overview of these arguments in section 6.2. Before I state my arguments, however, I need to provide the reader with some information about realization. We are already familiar with some of this background information from chapter 2. So, here I expose only the most important issues for my purposes.

6.1. Species of realization

First, the realization relation is considered by many to be a relation of ontological dependence (Gillett 2007a, p. 166). The received view amongst realization theorists, regardless of their different persuasions, is that this form of dependency is *synchronic and noncausal*. For instance, as Bennett, in her survey of metaphysical dependence relations, specifies: "Building relations do not unfold over time. If property *P* realizes property *Q*, it does so at some time *t* […]. Causation, by contrast, is paradigmatically *diachronic*, and that idea is frequently invoked to distinguish causation from relations like [realization]." (2011, pp. 93-94; italics in original)[51] Thus, realization and causation are distinct modes of determination, since one holds *diachronically* (causation), whereas the other holds *synchronically*, or at a durationless instant *t* (realization).

Second, consider the distinction between what Gillett calls the "flat" view, on the one hand, and the "dimensioned" view, on the other. What Gillett terms the flat view of realization is a one-one relation between properties and/or property instances instantiated by the same individual, and where the realizer/realized properties are individuated by their causal role (Gillett 2002, p. 317). Furthermore, the flat view – or, as Gillett also calls it, the standard view – is the conjunction of two metaphysical theses. The first thesis concerns *sufficiency* such that property *P* realizes property *Q* only if the causal powers of *Q* are a subset of *P*. Gillett calls this the *metaphysical sufficiency thesis*: "(1) Property instances P1-Pn are realizers of property instance F, at time t, *if and only if* P1-Pn are a minimal combination of

[51] Or, as Aizawa & Gillett put it: "[We] should mark that relations like realization are obviously a species of determination relation, but are different from causal relations. The 'horizontal' determination involved with causation is temporally extended, relates wholly distinct entities and often involves the transfer of energy and/or the mediation of force. In contrast, compositional relations are not temporal in nature, since their 'vertical' determination is synchronous, does not relate wholly distinct entities, and does not involve the transfer of energy and/or mediation of force. Composition is thus a variety of what has been termed 'non-causal' determination." (2009a, p. 198) Note that Aizawa & Gillett use the term 'composition' broadly so that it includes the realization relation.

property instances which are together metaphysically sufficient for an instance of F at t. (MS-thesis).” (2007a, p. 174; italics in original) Wilson also states that the standard view of realization takes the form of sufficiency: “realizers are metaphysically sufficient for the properties or states they realize.” (2001, p. 4)[52] The second thesis of the flat view concerns the individuals in which the realizer/realized properties are instantiated. Gillett frames this view accordingly: “A property instance X realizes a property instance Y *only if* X and Y are instantiated in the same individual.” (2002, p. 317; italics in original) Wilson dubs this for the constitutivity thesis of the standard view: “realizers of states and properties are exhaustively physically constituted by the intrinsic, physical states of the individual whose states or properties they are.” (2001, p. 5)[53] Gillett thinks that the flat view is inadequate, and offers, in its stead, a dimensioned view of realization. Advocates of the dimensioned view think that realization only trades in *properties* as relata[54]. Also, that the realizer/realized properties are usually qualitatively distinct. Take

---

[52] Gillett refers to both Kim (1998) and Shoemaker (2001) as proponents of the flat view of realization and, consequently, the MS-thesis. For instance, as Shoemaker states: “… property X realizes property Y just in case the conditional powers bestowed by Y are a subset of the conditional powers bestowed by X …” (2001, p. 78; cited in Gillett 2002, p. 318) And as Gillett frames Kim’s view: “A property instance X realizes a property instance Y *only if* the causal powers individuative of the instance of Y match causal powers contributed by the instance X (and where X may contribute powers not individuative of Y).” (2002, p. 318)

[53] Philosophers like Kim (1998) and Shoemaker (1999) express commitment to something like the constitutivity thesis. As Kim states: “It is evident that *a second-order property and its realizers are at the same level […] they are properties of the very same individual*.” (1998, p. 82; italics in original) Or, as Gillett (2002) frames Shoemaker’s view: “A property instance X realizes a property instance Y *only if* X and Y are instantiated in the same individual.” (2002, p. 317; see Shoemaker 1999, p. 297) As expressed in the constitutivity thesis, both Kim and Shoemaker demand that realizer/realized properties are contained in the same individual.

[54] Opponents of the dimensioned view such as Polger (2010) do not discriminate as to which kinds of entities – e.g., properties, processes, objects, and so on – are related by realization.

Gillett's example of a cut diamond, *S** (2002). *S** has the realized property of being very hard, *H*. Suppose *H* is composed by carbon atoms *S1-Sn*, and that *S1-Sn* have the properties of being bonded, *B1-Bn*, and being aligned, *A1-An*. *S** has *H* but has neither *A1-An* nor *B1-Bn*. Similarly for the carbon atoms, which have *A1-An* and *B1-Bn* but not *H*. In addition to these features, the dimensioned view takes realization to be a many-one relation rather than a one-one relation – e.g., many carbon atoms combine to compose one diamond. Moreover, in their recently offered accounts of realization, Gillett (2007a) and Aizawa & Gillett (2009a, 2009b) reject both the use of the constitutivity thesis and the metaphysical sufficiency thesis of the flat view in providing accounts of realization. These accounts reject the constitutivity thesis, since the realizer/realized properties may, in the right circumstances, be instantiated in *different individuals*, whereas the flat view presupposes that realized properties are exhaustively dependent on the internal properties of the individual whose properties they are (Wilson 2001, p. 4). As Gillett states this feature of the dimensioned view: "H is not identical to any of the particular properties/relations of any individual carbon atom, for H is instantiated in the diamond whilst particular relations of bonding and alignment are instantiated in some carbon atom." (2002, p. 319) We should also note that H comprises different *causal powers* than any of the causal powers instantiated by the properties of bonding and alignment of a carbon atom (Gillett 2002, p. 319). On the dimensioned view, then, parts and whole are *individuals*, which bear powerful causal relations to one another. The dimensioned view rejects the metaphysical sufficiency thesis of the flat view, since the dimensioned view is indexed against background conditions, while the sufficiency thesis presupposes the necessity of background conditions (Gillett 2007a). In this sense, realizers are *spatially contained* within the individual that is the composed entity.

The distinction between the flat view and the dimensioned view, together with the fact that realization is understood to be a relation of synchronic dependence, give us some idea for the different varieties of theories of realization. However, I want to finish this sketch of different species of realization by introducing the *wide realization view* proposed by Wilson (2001, 2004a, 2004b; Wilson & Clark 2009). According to Wilson, there are cases of the realizer/realized relation in which the realizers of some realized property P extend beyond the boundary of the individual bearer, IB, who has P. This characterization rejects the constitutivity thesis of the flat view. As Wilson states: "[…], wide realizations […] extend beyond the physical boundary of the individual, they are not exhaustively constituted by the intrinsic, physical properties of the individual subject, and so do not satisfy the constitutivity thesis." (2001, p. 12) However, the wide realization view accepts the metaphysical sufficiency thesis, in the sense that it is only the physical properties constituting a total realization *together with the appropriate background conditions* that metaphysically suffice for P. As an example of putatively wide realization, Wilson claims that: "Fitness is a dispositional property of individual organisms (or even whole species), this disposition is not individualistic, since physically identical organisms may differ in fitness because they have been or are located in different environments." (2001, p. 13) According to the wide realization view, then, in the right circumstances, only properties instantiated *within* the individual *together with* properties instantiated beyond that individual's brain and/or entire bodily constitution metaphysically suffice for some realized property.

6.2. Arguments

With this sketch of different species of realization in hand, I now turn to set up the arguments that I shall pursue in order to discuss the potential confluence between the metaphysics of realization and free energy minimization.

The first argument is the following: If the world is such that physical realizers are metaphysically sufficient for what they realize, and if the physical realizers are wholly instantiated within the same individual as the realized property, then the following view of realization holds: the relationship between realizer/realized cannot be such that the physical realizers can be exemplified in individuals different from the realized property. On a first pass, this argument turns on and, as a result, supports Gillett's critiques of the flat view of realization. That is, if the world is such – as purported within the free energy framework – that the property of free energy minimization and its realizer properties are instantiated within disparate individuals, then the following claim about realization holds: the relationship between realizer/realized properties may, in the right circumstances, be instantiated in different individuals. However, on a second pass, this argument rejects Gillett's presupposition that the components, whose properties enter into relations of realization, are *spatially contained* within the individual associated that is the composed entity. That is, if the world is as it is claimed to be in the free energy framework, then the systemic components and their properties that realize the property of free energy minimization may, in the right circumstances, outstrip the boundaries of the skin-and-skull to include properties of that individual's extra-neural and/or extra-bodily environment. If this is correct, it follows that free energy minimization is widely realized (Wilson 2001).

To underpin both of these claims, consider that on its *thermodynamic* axis, the free energy principle starts from the premise that biological systems are dissipative systems. A dissipative system is an open system, which operates far-from-thermodynamic equilibrium by exchanging energy or entropy with the surrounding environment (Friston & Stephan 2007). Here I should note that according to the free energy framework, biological systems preserve their order, despite being immersed in an environment that is irrevocably becoming more disordered, since the environment "unfolds in a thermodynamically structured and lawful way and biological systems embed these laws into their anatomy." (Friston & Stephan 2007, p. 422) As a result, the realizers of entropy minimization include, necessarily, properties of that individual's extra-neural and extra-bodily environment. If we maintained the internal structure associated with (or which is) the individual composed, whose realized property of free energy minimization it is, but varied the nature of the extra-neural and/or extra-bodily environment, the realized property would alter radically.

I will provide a similar argument as I consider the *information-theoretic* axis of the free energy principle. If manipulation of environmental structures, embedded in continuous loops of perception and action, affords extra-bodily circuitry for the minimization of prediction error, then this violates the view that the physical realizers of P (that is, the minimization of prediction error) must be spatially contained within an individual S. Hence, if the world is as suggested by the free energy principle, then the realization base of P must be wide. We thus have two arguments: one that contravenes the flat view, in the sense that the realizer/realized properties need not be exemplified within one individual; and another argument that contravenes the dimensioned view: that the components, whose properties enter into relations of realization, are spatially contained within the individual associated with (or which is) the composed entity.

The second argument I shall examine turns on the idea that the received view of realization is a synchronic (and noncausal) relation. The form of the argument is conditional. Irrespective of one's preferred version of realization, if the world is such that realization and free energy minimization can be brought together, then it follows that: the relationship between realizer/realized is such that there is a synchronic relation between the realized properties and their physical realizers. As we have seen, the presupposition that realization is a synchronic (durationless or atemporal) relation of dependence is a commonly accepted condition on realization – one endorsed by defenders of the flat view, the dimensioned view, and the wide view. However, if the world is such – as stated by the free energy principle – that free energy minimization is a property of dynamical systems, whose components are orchestrated in temporally extended processes with properties such as nonlinearity, then the following claim holds: free energy minimization is a property of temporally unfolding processes, and the latter are themselves composed of temporally unfolding processes. If this is true, then the synchronic conception of realization is problematic.

A clue to the needed account: the difference between objects and their properties, on the one hand, and processes and their properties, on the other. Objects, such as a chair, and its properties, such as being brown, are *timeless*, in the sense that they exist whole and complete at a synchronic instant *t*. A diamond's property of being hard is wholly present at each moment in time during which the diamond persists. In this sense, objects or object-like entities have no temporal extent (Noë 2006). But, processes, and their properties, are temporally extended in nature. To suppose that the property of free energy minimization exemplified in dynamical systems is complete at a synchronic moment in time would be to confuse the relation between processes and their properties with the relationship between object-like entities and their properties.

In addition to this, consider that a common strategy by which to identify what "constitutes" the realization base of some realized property, is by appealing to what plays the most salient causal role(s) in relation to the instantiation of that realized property (Cosmelli & Thompson 2010, p. 364). But, if the systems that minimize free energy are complex (nonlinear, self-organizing, and temporally dynamic) systems – as argued by proponents of the free energy principle (Friston & Stephan 2007) – then the question of what plays the most salient role will be difficult to answer. The reason for this is that the behavior of single neurons and/or neuronal assemblies cannot be determined independently of global brain activity and the timescales over which such activity unfolds. As Cosmelli & Thompson explain: "In dense nonlinear systems in which all state variables interact with each other, any change in an individual variable becomes inseparable from the state of the rest of the system." (2010, p. 365) Moreover, in nonlinear systems, individual neurons or neuronal assemblies operate over multiple different time-courses, even though the activity of different neurons or neuronal assemblies may operate synchronously (Friston & Stephan 2007; Varela et al. 2001). Thus, if the world is as suggested by the free energy principle, it follows that the minimization of free energy refuses realization at a synchronic instant. Perhaps this shows that the realization relation is apt for a diachronic-friendly extension. That is, could we not simply say that such-and-such temporal features realize free energy minimization? In this chapter, I will not argue for this point; rather, my aim is to show that the synchronic constraint of the received views on realization is ill fitted to do the job.

6.3. Overview

In section 6.4, I deal with the first argument outlined in section 6.2. In so doing, I will problematize the adequacy of both the flat view and the dimensioned view. In section 6.5, I develop the second argument outlined in section 6.2, where I will scrutinize the synchronicity constraint of the usual theories on realization. I conclude this chapter differently than the rest of the chapters, in the sense that the best I can hope to show is that given certain conditions, it is (arguably) the case that free energy minimization cannot be brought together with the realization relation.

6.4. Argument #1: Realization, wide realization, and the free energy principle

6.4.1. Argument from thermodynamics

Recall that the flat view of realization states that the world is such that the realizers are metaphysically sufficient for what they realize, and that the physical realizers are instantiated in the same individual as the realized property. If true, the relationship between realizer/realized properties cannot be such that physical realizers are instantiated in individuals different from that of the realized property, and the realizers are, necessarily, sufficient for the exemplification of the realized property. However, according to the free energy minimization formulation, the world is not like this. If the world is as the free energy principle claims, the relation between realizer/realized properties is such that these properties are, in the right circumstances, instantiated in different individuals.

Consider that the free energy principle states that all physical systems (in order to survive) must actively resist a natural tendency for disorder (Friston 2003, 2010, 2011; see also Ashby (1952) and Haken (1983)). This is the thermodynamic starting point of the free energy principle, and it brings the free energy principle into alignment with principles of dynamical systems, the central premise of which is that physical systems, in general, and biological systems, in particular, belong "to a class of systems that are both *complex* and that exist *far from thermodynamic equilibrium*." (Thelen & Smith 1994, p. 51; italics in original) Biological systems are complex, in the sense that such systems typically consist of many components, and these components tend to be different with disparate properties as well as causal powers. In addition, biological systems exist far-from-thermodynamic equilibrium, because such systems contravene the second law of thermodynamics. The second law of thermodynamics states that entropy (i.e., a measure of disorder, or, more simply, the number of way the components of a system can be rearranged) of closed systems increases over time (Friston 2010, p. 127).

We know that organisms are capable of maintaining reduced levels of entropy in the face of fluctuations and increasing levels of entropic disorder in the external environment. In his discussion of which kinds of properties and components must be involved in realizing such a capacity, Kemp (1982) mentions, among other things, the role of temperature regulation to sustain appropriate levels of internal temperature. But this is not possible without some sort of blood filter (i.e., circulatory system), the property of which is to filtrate and pump blood throughout the body. We do not need to add additional components and properties to this example in order to establish the following: that the capacity of organisms to maintain entropy is realized in a number of different components. If Gillett (2002, 2007a) is correct to insist that components are individuals, then it follows that the capacity of entropy minimization is a property of organisms realized by different individuals. Here I have highlighted intralevel components, i.e., constituent components. But we should note that the same point holds between systemic levels. Although the microphysical parts involved in realizing entropy minimization may be qualitatively different, they combine to compose the qualitatively different property of entropy minimization.

In this case, we have an argument against the assumption of the flat view, that the realization relation is a one-one relation that holds within one and only one individual. As a first approximation, this supports the dimensioned view. First approximations, however, are not always correct. This is one of those occasions. That is, I shall now show that Gillett's presupposition that the components, whose properties enter into relations of realization, are spatially contained within the individual associated with (or which is) the composed entity is problematic – at least in the case of free energy minimization. If the world is as stated by the free energy principle, then the systemic parts and their properties that putatively realize free energy minimization are not spatially contained within an individual but includes components and properties of that individual's extra-neural and/or extra-bodily environment.

Consider that in contrast to closed systems, biological systems are open systems. Open systems are also called dissipative systems, which is meant to specify that such systems preserve their order, while being immersed in a dynamical environment, by exchanging energy or matter with that environment. Using this terminology, we can state that the amount of entropy in a system is negatively correlated with its potential for survival (Friston & Stephan 2007, p. 423). In short, for biological systems to avoid *phase-transition* (e.g., from being alive to being dead), they must exchange entropy with the environment.

Combining the thermodynamics of free energy minimization with the realization relation, let us start with the following. Call H the property of self-maintenance, X the property of drawing energy from the environment, P the property of manipulating an energy source, and R for the property of dissipation. Here X can at best be a *partial* realization of H – and similarly for P and R. A partial realization is what Shoemaker (1981) calls a *core* realization, which is a particular component of the central nervous system, say, that is identifiable as performing a core role in bringing about H. If this turns out to be correct, which is the norm in the literature, then partial realizations alone will not satisfy as metaphysically sufficient for H. This does not yet provide us with an argument against the flat view. But that core realizers are insufficient for H identifies the need for something extra. That is precisely what Shoemaker (1981) provides. According to Shoemaker, when considering the relation between some realized state or process (e.g., H), and the system, S, in which H is realized, one must distinguish between *core* realizations and *total* realizations. In his discussion of Shoemaker's account, Wilson provides the following definition of core- and total-realization (2001, p. 8; italics in original):

(a) *Core* realization of H: a state of the specific part of S that is most readily identifiable as playing a crucial role in producing or sustaining H.

(b) *Total* realization of H: a state of S, containing any given core realization as a proper part, that is metaphysically sufficient for H.

Wilson does not discuss the case of self-maintenance, even though he uses the placeholder H in his definitions. With this clarified, consider that total realizations of H (that is, the property of self-maintenance) includes X, P, and R. In this sense, total realizations are *complete* realizations. However, if the world is as stated by the free energy principle, then even if total realizations are complete, they are still metaphysically insufficient for H, since – as Wilson would say – the total realization of H "excludes the *background conditions* that are necessary for there to be the appropriate, functioning system." (2001, p. 9; italics in original)

If we consider the thermodynamic formulation of the free energy principle, it becomes apparent why it is only the physical states that make up the total realization in conjunction with appropriate extra-bodily properties that will metaphysically suffice for realizing H. A *total* realization of H includes X, P, and R. However, excluded from the total realization of H is the necessary fact that the environment itself "unfolds in a thermodynamically structured and lawful way […]" (Friston & Stephan 2007, p. 422), which is necessary for the system S to function the way it does. Strictly speaking, for an open system to maintain its structure and function it must engage in energy exchange with the environment. But, for this to be possible, the environment itself must be able to continuously consume and dissipate energy[55].

[55] Shapiro points to one aspect of this in his discussion of homeostasis as a constraint upon biological systems. For instance, noting that the external

Call these extra-bodily or environmental realizers of H, ER. Importantly, ER's are not part of the total realizations of H, instantiated in S, since the ER's are not properties of S – the individual within which the total realizations of H are spatially contained. Thus, if H is a *realized* property, the properties realizing H are not wholly spatially contained within the individual S, where S is the individual instantiating H. This fact counts against the dimensioned view of realization.

There is at least one reason to believe that this outcome (against the dimensioned view) is premature. For example, Gillett (2007a) provides a critique of the appeal to "external" entities as actual physical realizers due to the fact that such an appeal turns on the view that such "external" entities are elements of a metaphysical sufficiency condition for the realized entity. On Gillett's view, however, metaphysical sufficiency "leads to scientific hyper-extension by placing realizers, and parts, beyond the normal scientific limits and understanding." (2007a, p. 176) Gillett argues, that in scientific examples, realizers are not metaphysically sufficient for realized properties, because – strictly speaking – only realizers together with entities that function as background conditions are sufficient for the realized properties. Gillett's view is based on what he terms "[our] well-confirmed scientific theories," (2007a, p. 176) which he finds in his analysis of examples from chemistry and biology.

Even if Gillett is correct in what he takes to be our well-formed scientific theories in chemistry and biology, it is important to mention that the free energy minimization formulation is "premised [on the fact] that the environment unfolds in a thermodynamically structured and lawful way and that biological systems embed these laws into their anatomy." (Friston & Stephan 2007, p. 422) Here we have a scientific theory that takes it as an integral fact that certain non-neural and extra-bodily properties of the environment – that environmental order is assured in the face of irreversible disorder by the fact that the environment is thermodynamically structured – are an inherent part of a physical organism's capacity to reverse an increase in physical entropy over time. As a result, and viewed under counterfactual

---

environment is less dilute than an organism's internal environment, the external environment enables an organism to regulate its temperature by 'pulling' water from the organism, "drying it out and altering its concentration of ions." (2004, p. 88) conditions, if we maintain the "internal" realizers but varied the nature of the environment such that it is not thermodynamically structured, then this would radically change the capacity of organisms to preserve their order.

It is also worth stressing that Gillett's appeal to *spatial containment*, as he specifies that components and their properties be spatially contained within the individual that is associated with (or which is) the entity composed (2007a, p. 166), does not find any corresponding image in contemporary physics (Ross & Ladyman 2010, p. 159). As Ross & Ladyman state: "The types of particles which physical theory describes do not have spatiotemporal boundaries in anything like what common sense takes for granted in conceptualizing everyday objects, and in that respect are not classical individuals – the philosopher's little things (French & Krause 2006)." (2010, p. 156) So, with the insights from the thermodynamic rendition of the free energy principle, we might even question the spatial containment condition of the dimensioned view. I do not think that this provides trouble free evidence for the claim that if the world is as purported by the free energy formulation, the realization relation must be *wide*. But, what it does establish is potential limitations with the dimensioned view of realization, while lending additional plausibility to a wide view of realization.

6.4.2. Introducing predictive processing in the free energy principle

Similar problems with both the flat view and the dimensioned view of realization arise when we consider the free energy principle from its information-theoretic perspective. Before arguing for this, I need to introduce several core aspects of the free energy principle in cognitive neuroscience. I begin exposing the relationship between free energy and Bayesian inference.

According to Friston et al., free energy "bounds surprise, conceived as the difference between an organism's predictions about its sensory inputs […] and the sensations it actually encounters." (2012, p. 1) "Surprise," in this context, does not refer to personal-level or conscious surprise (e.g., the kind of surprise one experiences when somebody organizes a surprise party for you). Instead, "surprise" in the free energy principle is understood as a measure of improbability from information theory. Similarly to the thermodynamics of the free energy principle, the information-theoretic formulation states: "organisms that succeed […] do so by minimizing their tendency to enter into this special kind of surprising (that is, non-anticipated) state." (Friston et al. 2012, p. 1) That is: "For the brain to be energetically efficient and for our behavior to be optimal and adaptive, we [our brain] utilize knowledge from previous experiences to make predictions about the future and minimize the cost of surprise." (Brown & Brüne 2012, p. 1)

Free energy minimization is modeled in the framework of Bayesian inference. As Brown & Brüne state: "Bayesian statistical inference is a mathematical method of inference which incorporates priors, or prior beliefs learned from previous experiences that generate internal models [i.e., probabilistic representations] of a predicted outcome, and consequently acts as top-down modulators of bottom-up sensory input." (2012, p. 3) According to Hohwy, Roepstorff, and Friston (2008), the predictive brain hypothesis explains how the brain, through the implementation of Bayesian inference, utilizes probabilistic representations of the causes of its sensory inputs to anticipate future events. The brain does this by making operative processes the function of which are to optimize a particular kind of *prediction error*, typically associated with the activity of superficial pyramidal cells; these cells being the source of the forward and backward connections in the brain (Brown et al. 2011, p. 2). In the mammalian brain, prediction error is corrected for within a hierarchy or cascade of cortico-thalamic and cortico-cortical processing in which ensembles of neurons attempt to predict the input generated at lower levels of computation in the architecture on the basis of their own probabilistic representations of the sensory input (Mumford 1992, p. 241). The predictions with the highest *posterior probability* (i.e., most probable given the input) fix the content of the sensory input (Hohwy et al. 2008, p. 688). If this picture of the brain's information-theoretic processing turns out to be correct, it is evidence for the fact that the brain (in Bayesian terms) makes use of a *generative model*, composed of two elements: *likelihood* (i.e., the probability of sensory input given their causes) and *prior probability* (i.e., how probable the prediction was before the input) (see e.g., Friston 2002; Hohwy et al. 2008).

Another feature of the predictive brain hypothesis is the idea that the brain's processing architecture is hierarchical. Hierarchical organization presupposes a distinction between forward and backward connections. Forward (bottom-up) connections run from lower to higher areas in the brain, while backward (top-down) connections go from higher to lower brain regions. Within a hierarchical level, lateral connections connect various other regions (Hohwy 2007; Hohwy et al. 2008). Neuroimaging studies suggest that forward connections carry sensory input, while backward connections modulate bottom-up inputs, *viz*., have a controlling influence over "the probability of certain aspects" of the content transmitted by sensory receptive fields (Friston 2003, p. 1328). To successfully suppress error signals, leaving only the prediction error to be passed forward in the system, the activity of *explaining away* becomes important. Explaining away involves a process of matching the driving sensory signal with a cascade of predictions emerging in spatiotemporal activity in the hierarchy. As Clark mentions: "Perception here becomes 'theory-laden' in at least one (rather specific) sense: what we perceive depends heavily upon the set of priors (including any relevant hyperpriors) that the brain brings to bear in its best attempt to predict the current signal." (2013, p. 187) Consequently, the brain (rather than being passive) is actively and continuously trying to predict the posterior probability – given the recognition density – at each processing level of the hierarchy.

6.4.2.1. Argument from predictive processing in the brain

The argument against the flat view above presupposed a particular reading, namely that realizers are wholly constituted by "internal" properties of individuals. Here I wish to take literally when Wilson specifies that part of the so-called flat view is that "realizers of states and properties are exhaustively constituted by the *intrinsic*, physical states of the individual whose states or properties they are." (2001, p. 5; italics added) "Intrinsic properties" are usually understood to be those properties that an entity possesses independently of everything else that exists or independently of there existing anything else (Ladyman & Ross 2007, p. 135).

If the world is as the flat view takes it to be, then it follows that realized properties are realized by the intrinsic physical properties of the individual bearer, whose properties they are. But, if the world is as stated in the free energy principle, realized properties prevent realization by intrinsic physical properties, because the properties of the realizers are themselves non-intrinsic.

Consider, firstly, what Rao & Ballard say about the bidirectional connectivity in predictive hierarchical architectures: "[Prediction] and error-correction cycles occur concurrently throughout the hierarchy, so top-down information influences lower-level estimates, and bottom-up information influences higher-level estimates of the input signal." (1999, p. 80) Or, as Varela et al., in their review of functional integration in the brain, say: "With only a few exceptions, the brain is organized on the basis of what we can call the principle of reciprocity: if area A connects to area B, then there are reciprocal connections from B to A." (2001, p. 230) To get a grip on this idea, consider, secondly, how Hohwy et al. (2008) explain the phenomenon of binocular rivalry.

Binocular rivalry is a form of subjective visual experience that occurs, in a special experimental setup, when one stimulus is shown to one eye and another stimulus is shown to the other. For example, when an image of a house is presented to the right eye, and an image of a face to the left eye, the subjective experience tends to unfold in a bi-stable manner by alternating between the house and the face. This is what is known as binocular rivalry. As Hohwy et al. explain, to account for binocular rivalry, two parts need explanation: first, the *selection problem*: "why is there a perceptual decision to select one stimulus for perception rather than the other, and, further, why is one of the two stimuli selected rather than some conjunction or blend of them?" (2008, p. 690), and second, the *alternation problem*: "why does perceptual inference alternative between the two stimuli rather than stick with the selected one?" (2008, p. 690)

From the perspective of Bayesian inference, if a subject is currently experiencing an image of a face, F, why, then, does the F hypothesis have the highest probability, given that F and house H have an equal likelihood? This is the selection problem. The alternation problem is to explain why the system (the brain), having selected F, say, after only a few seconds de-selects in favor of H. Note that for my present purposes, discussion of both the selection and the alternation problem is unnecessary, so I shall restrict my attention to the alternation problem here. According to Hohwy et al., the predictive processing framework posits a hierarchical inversion of generative models of how inputs are caused to explain the alternation problem:

> "At the higher, hypothesis-generating level only the currently best hypothesis is allowed to generate predictions. It seems plausible that inhibition will be lateral, in relation to other hypotheses at the same level. This gives high activity for the winning hypothesis with the highest posterior and this for the

> dominant percept, and lower activity for other hypotheses at that level. At the lower level there is the opposite pattern: the bottom-up driving signal for the dominating percept is explained away by good predictions, meaning the prediction error for the dominant hypothesis is suppressed. Conversely, the bottom-up error signal for the currently suppressed stimulus is not.” (2008, p. 691)

As with the property of *self-maintenance*, free energy minimization is ineliminably relational, with all areas (thalamo-cortical, cortico-cortical) working simultaneously, yet at different temporal frequencies. Order is maintained in the overall processing, Friston & Stephan explain, through *synchronous* activity in the various top-down and bottom-up loops in the computational architecture (2007, p. 443).

In dichoptic viewing conditions, when F is viewed by one eye and H by the other one, the hypothesis with the highest *prior probability* (how probable the prediction was before the input) could be considered as a core realization, *viz*., a specific part of S that is identifiable as playing a crucial role in producing the realized property. But, selected hypotheses are metaphysically context-sensitive, in the sense that they will realize a visual experience of F, say, only in relation to their activity and location within a generative hierarchical organization (Friston 2002).

While total realizations are said to be complete realizations, the assumption of the flat view that realized properties are realized by physical intrinsic realizers is inconsistent with the free energy principle, because a crucial property of the process of functional integration – in the processing architecture – is *temporal synchrony* (Engel 2010; Engel et al. 2001; Friston 2003; Varela et al. 2001). In contrast to the idea that intrinsic realizers realize properties, temporally synchronous patterns are extrinsic, relational properties of dynamical systems such as the brain. As a result, if the free energy principle is correct, then the flat view of realization is false, since the physical realizers are themselves (in part, at least) relational (i.e., non-intrinsic).

6.4.2.2. Argument from wide predictive processing

What I call the argument from wide predictive processing is intended to show that the spatial containment constraint of the dimensioned view is problematic. The intended argument is as follows: if active manipulation of worldly resources, embedded in on-going loops of perception and action, afford extra-bodily circuitry for minimization of prediction error, then the spatial containment condition of the dimensioned view is at best unproven in the case of free energy minimization. This lends support (*prima facie*, at least) to a wide conception of realization.

In their "Enculturating brains through patterned practices," Roepstorff et al. (2010) go on to suggest that the brain is a hierarchically organized predictive machine, which attempts to anticipate its sensory inputs based on empirical priors of causes in the environment. That idea finds its fullest expression in the *patterned practice approach* in social anthropology and social neuroscience. In Roepstorff et al. words:

> "The patterned practice approach is highly compatible with these findings [predictive brain hypothesis]. […]. The affinities between 'predictive brain' models and a patterned practice approach may not be merely metaphorical. At different levels, they frame the link between action and perception as a continuous process of resonance, where networks-in-action order the coordination of input and output as networks-in-action form and unfold in practice." (2010, pp. 1056-57)

The idea is that just as top-down predictions modulate bottom-up input so can socially embedded and culturally transmitted practices be understood as modulatory. Evidence for this is provided by Roepstorff & Frith (2004), who consider the concept of "top-top" modulatory control of action in a study of the 'Wisconsin card-sorting task' (WCST). Based on brain imaging experiments, Roepstorff & Frith (2004) argue that the state-oriented (here-and-now time perspective) "top" in "top-down" driving and modulatory control of action rather than being conceived as *internal* to the experimental participant is in fact socially distributed across the experimenter and experimental participant in cognitive experiments. Roepstorff & Frith focus on several experiments. I shall focus on one of these, namely a study on the cross-species neural correlates of action conducted by Nakahara et al. (2002).

Nakahara et al. (2002) had two macaque monkeys perform a version of the WCST. In a WCST, which consists of four cards and 128 response cards with geometric figures that vary according to perceptual dimensions such as color, form or number, the experimental participant is presented with cards that display specific symbols of one of the three perceptual dimensions, such as three green circles, or three yellow triangles, etc. The task requires the participant to find the correct classification rule, *viz*., sorting criteria. During the task, the participant is given feedback related to the correctness of their sort. Once the participant chooses the correct rule they must maintain the use of this rule irrespective of the fact that the stimulus changes. After a certain number of correct matches, the experimenter changes the sorting criteria without warning, demanding the participant to discover the new classification rule.

During the task, Nakahara et al. had the two monkeys perform a computerized version of the WCST, where the monkeys had to select one of three cards relative to the classification rule in use at the time of sorting. The feedback was provided visually on the screen and the monkeys received liquid as a reward for choosing correct. The MRI results showed that "the main effect of the set-shifting component of the WCST was found in the ventrolateral prefrontal cortex bilaterally, […] at the ventral end of the inferior ramus of the arcuate sulcus." (Roepstorff & Frith 2004, p. 191) Nakahara et al. also had 10 human subjects perform the same task. As Roepstorff & Frith specify: "In these [human] participants, the main activation was found in the posterior part of the bilateral inferior frontal sulcus (Broadmann's area 44/45, […])." (2004, p. 191) These results, both Nakahara et al. (2002) and Roepstorff & Frith (2004) agree, confirm that the main sites of activation in the two species may be considered functionally and anatomically homologous.

To perform well in the WCST, the participant must enact and modify a particular cognitive set, "which can be used as a template for acting in the world […]." (Roepstorff & Frith 2004, p. 191) According to Roepstorff & Frith, this is a clear indication of a top-down control of action in both an anatomical sense (from prefrontal to lower brain areas) and in the predictive processing sense (denoting a form of hypothesis or prediction-driven processing). Adding to the central conclusion by Nakahara et al., that there is evidence of cross-species neural correlates of action, Roepstorff & Frith provide an alternative interpretation of the experimental outcomes, one that is based on a combination of patterned practices and the different developmental and social trajectories between the macaque monkeys, on the one hand, and the humans, on the other[56].

The important thing to note is that whereas it took Nakahara et al. up to one full year of training to get the monkeys to perform the WCST in the MRI scanner, it only took the human participants 30-60 minutes of verbal instruction to perform equally well. Thus, despite displaying similar patterns of behavior and brain activation, the learning trajectory between the two species is quite different. In the human case, Roepstorff & Frith stress, the internal top-down story breaks down. On the standard view, bottom-up effects are driven through sensory inputs established "from the outside," whereas top-down predictions are generated "from the inside," e.g., via knowledge-driven predictions about the causes of the sensorium. But, as Roepstorff & Frith argue:

> "[The] 'verbal instructions' that enable the human volunteers to perform well in the task, fail to fit this scheme. The instructions are clearly coming 'from the outside' and are mediated via the senses, i.e., bottom up, and yet their main purpose is to allow for the very rapid establishment of a consistent model of how the participants are to interpret and respond in the situation, i.e., top-down." (2004, p. 192)

The main result that Roepstorff & Frith point to is that given this breakdown of the conventional model of the "top" in top-down processing, "the origin of the 'executive top' employed in the WCST is out-side the brain of the participant, namely [socially mediated by the] experimenter." (2004, p. 194) They refer to this

[56] Roepstorff et al., define their notion of "patterned practice" as follows: "Everyday life is continuously ordered into more or less stable patterns that are specific to particular types of situations, defining preferences, predispositions, and expectations for actors. […]. These patterns present regularities that arise from everyday practices while at the same time shaping them." (2010, p. 1051) The idea behind the patterned practice approach is to highlight the following: "A patterned practice approach assumes that regular, patterned activities shape the human mind and body through embodiment, and internalization. Vice versa, enacting practices shape and re-shape norms, processes, institutions, and forms of sociality." (2010, p. 1052)

socially mediated form of interaction between the experimenter and the participant as a “top-top exchange of scripts.” (2004, p. 192) Scripts are “shared representations” enacted in situated practices, where shared representations concern top-level aspects of control (that is, the goal of the task) instead of low-level processes concerning (e.g., how specific movements should be made). According to Roepstorff et al:

> “From the inside of a practice, certain models [i.e., certain ways of interacting with one another] of expectancy come to be established, and the patterns, which over time emerge from these practices, guide perception as well as action.” (2010, p. 1056)

As with predictive processing in the brain, one property of dynamical processes that regulates the coherency and resonance between patterns of expectancy in the brain and patterns of expectancy unfolding in the social context is *temporal dynamics*. For instance, forward neural connections mediate their post-synaptic effects over very fast timescales, ranging from 1.5-6 ms decay time, while backward neural connections are mediated by slower dynamics, with ～ 50 ms decay time (*cf.* Friston 2003, p. 1328).

According to Friston, slower neural dynamics mediate contextually enduring effects, which is why backward neural connections can modulate forward neural connections. This difference between forward and backward neural connections, Friston refers to as “functional asymmetry,” to emphasize the difference in functional role between those neural connections. In the WCST case, the proposal is (among other things) that top-down predictions – in the context of culturally mediated practices – take the form of socially situated top-top interaction in patterned practices. That is, the interaction between experimenter and participant display temporal dynamics that are much slower (ranging from 30-60 minutes) than forward and backward neural connections. If slower evolving dynamics mediated contextual relevant information, it is probable that certain situated practices may display modulatory effects. As Roepstorff et al. mention, from the inside of a practice, certain expectancies and regularities emerge – and these expectancies and regularities may guide perception and action, and consequently minimize prediction error.

Recall that the dimensioned view of realization presupposes that components and their properties are spatially contained within the individual associated with (or which is) the composed entity. As a result, the realizer/realized properties are also spatially contained with the individual associated with (or which is) the composed entity. On Gillett's view, if we treat cultural practices and other extra-bodily components and properties as *physical realizers* of some property, we fail to discriminate between physical realizers and the *background conditions* (i.e., causal conditions) necessary for those physical realizers (2007a, p. 175). But what if certain properties of computational processing simply could not be realized in the absence of particular ways of *being in the world*, then we would have reason to believe that specific situations and cultural practices are not merely causally enabling or necessary background conditions for predictive processing, but also realizers. This is the impetus behind Wilson's account of *wide realization* (2001; Wilson & Clark 2009).

We usually distinguish between realizers and background conditions through an analysis of which parts and other components are plausibly entities that 'play the most salient causal role' or simply 'play the role' of the composed entity. In other words, realizers are entities whose productive 'causal function' results in productive 'causal functions' of the composed entity. By contrast, entities that are merely background conditions do not 'play the role' of the composed entity.

From a patterned practice approach in social cognitive neuroscience (Roepstorff & Frith 2004; Roepstorff et al. 2010), it would seem that the patterned interactions between experimenter and participant could not simply be screened off as background conditions for predictive processing. As Roepstorff et al. stress: "From the inside of a practice, certain models of expectancy come to be established, and the patterns, which over time emerge from these practices, guide perception as well as action." (2010, p. 1056) If that is correct, then those *patterns of expectancy* play *parts of* the role of the composed entity: predictive processing. Thus, insofar as the patterned practice approach is true, it lends support to the claim that certain instances of free energy minimization have a wide realization base.

6.5. Argument #2: Realization, synchronicity, and the free energy principle

Thus far I have argued that there are problems with both the flat view of realization and the dimensioned view. The account of realization that has seemed most promising so far has been the wide realization view.

I now turn to consider the second argument of the chapter: if the world is such – as stated by the free energy principle – that free energy minimization is a property of dynamical systems, whose components are orchestrated in temporally extended processes with properties such as nonlinearity, then the following claim holds: free energy minimization is a property of temporally unfolding processes, and the latter are themselves composed of temporally unfolding processes. If this is true, then the synchronic conception of realization is problematic.

The starting point for the argument is that it is coherent to distinguish between a conceptual argument for realization and an empirical argument for realization. That is, if there is a relation of realization between realizer/realized properties, then that relation must hold synchronically by definition. This is a conceptual argument for the claim that realization is synchronic. We can outline the argument as follows: (i) the synchronic nature of realization serves to distinguish it from causation; (ii) causation is a diachronic relation; (iii) therefore, realization is not a diachronic relation; and (iv), therefore, realization is a synchronic relation. The inference from the premises (i) and (ii) to the conclusion (iii), and the consequent conclusion (iv), is valid.

The problem with the argument is with the evidence for premise (i): that the synchronic nature of realization serves to distinguish it from the relation of causation. It is at this stage in the argument that things become much more uncertain, but, I submit, also much more interesting. That is, if the world is as stated by the free energy minimization formulation, *viz*., that processes and their properties involved in the minimization of free energy are embedded in the temporal and nonlinear patterns of top-down and bottom-up computational processing, which itself is embedded in patterns of temporal and nonlinear activity, then the following would be the case: free energy minimization is a property of dynamical spatiotemporal processes or patterns, which themselves are composed of spatiotemporal processes or patterns of activity, none of which can be completely or wholly present at any particular moment in time (see e.g., Kelso 1995).

Suppose we agree (as seems empirically plausible) that dynamical systems such as the brain fail to instantiate the property of free energy minimization without that property being embedded in temporally unfolding and integrated ensembles of neuronal assemblies over time (Engel et al. 2001; Engel 2010). We thus agree that for the exemplification of free energy minimization to materialize, this necessitates the unfolding of highly complex and global temporal dynamics in the brain. That is, the rates of change within, the time-course of, and any time-dependent synchronization of individual neurons or assemblies of neurons (or groups of neurons) are non-trivially part of the realization of free energy minimization. If the free energy picture turns out to be correct, and due to the inherent need for temporal unfolding for free energy minimization to occur, it would seem that insofar as there is a realization base for free energy minimization that this realization base cannot determine the occurrence of free energy minimization synchronically. Consider, e.g., the following illustration in Friston & Stephan (2007) depicted the quantities that define free energy:

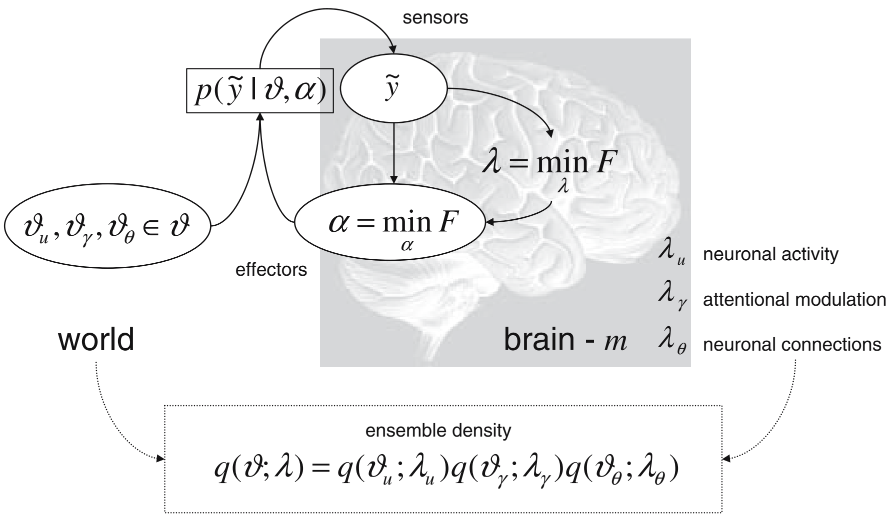

Fig. 11 Illustration detailing the quantities that define free energy (Friston & Stephan 2007, p. 424)

This illustration informs us of two things. First, it describes the quantities of a system's (*m*) interaction (or exchange) with the environment. Second, it indicates that all the quantities that change do so in order to minimize free energy. The latter is shown in the formulation λ = min *F* (λ refers to systemic quantities, whereas *F* denotes free energy). ŷ characterizes effects of the environment on the system; *a* is the effects of the system on the environment. According to Friston & Stephan, biological systems can minimize free energy by changing the two quantities that free energy depends on: (i) a system can act on the environment (*a*), thus changing the sensory input *ŷ*; (ii) or, a system can change its recognition density by changing its internal states, λ. The first path by which biological systems can minimize free energy can be represented as a conditional probability formulation of the form $p(\hat{y}) \rightarrow p(\hat{y} \mid a)$. In general, embodied manipulation of the environment can change the sensory input of *m* given the general form of the probability condition, i.e., the probability of an effect *e* occurring given that *f* occurs. Simply put, a system can minimize free energy by acting on the world, thus optimizing the accuracy of its own predictions by actively sampling and sculpting the environment. More formally "($p(\hat{y} \mid \upsilon, a)$ is the conditional probability of sensory input given its causes, υ, and the state of its effectors (i.e., action)." (Friston & Stephan 2007, p. 424) The remaining pathway by which a system can minimize free energy is by changing its internal states (λ). While portrayed as analytically separable, both pathways have functionally and structurally convergent dynamics. As Friston notes: "Internal brain states and action minimize free energy […], which is a function of sensory input and a probabilistic representation […] of its causes." (2010, p. 128) Thus, all the quantities that can change do so to minimize free energy.

One interpretation of figure 11, although arguably an incorrect interpretation, is that free energy minimization is physically realized by quantities $q(\upsilon, \lambda) = q(\upsilon_u; \lambda_u)q(\upsilon_y; \lambda_y)q(\upsilon_{,\theta}; \lambda_\theta)$ such that free energy minimization is simultaneously present with its realizers. The assumption that free energy minimization is realized at a single or durationless moment in time is an assumption that turns, I suspect, on our tendency to represent it spatially or pictorially as in figure 11. But the spatial representation is misleading, in the sense that the spatial (inert) representation is not analogous with the temporal dynamics through which free energy minimization is realized.

We already know that all quantities involved in free energy minimization change to minimize free energy. Consider that the quantities describing both environmental (or, hidden) causes, *υ*, and quantities describing neural states unfold and change on a timescale of milliseconds, seconds, and minutes. As Friston &

Stephan remind us, environmental causes could be large and heterogeneous in number. As they point out, a “key difference among them is the timescales over which they change.” (2007, p. 429) In figure 11, these environmental causes are partitioned into three sets, $\upsilon = \upsilon_u, \upsilon_y, \upsilon_\theta$, indicating change on a timescale of milliseconds, seconds, and minutes. According to Friston & Stephan, the “induces a partitioning of the system’s parameters into $\lambda = \lambda_u, \lambda_y, \lambda_\theta$ that encode time-varying marginals of the ensemble density.” (2007, p. 429) As they specify:

> “The first, λ_u, are system quantities that change rapidly. These could correspond to neuronal activity or electromagnetic states of the brain that change with a timescale of milliseconds. The causes υ_u they encode correspond to evolving environmental states, for example, changes in the environment caused by structural instabilities or other organisms. The second partition λ_y changes more slowly, over seconds. These could correspond to the kinetics of molecular signaling in neurons; for example calcium-dependent mechanisms underlying short-term changes in synaptic efficacy and classical neuromodulatory effects. […]. Finally, λ_θ represent system quantities that change slowly; for example long-term changes in synaptic connections during experience-dependent plasticity, or the deployment of axons that change on a neurodevelopmental timescale.” (2007, p. 429)

All of these quantities, then, are part of the physical machinery determining free energy minimization, and all of these quantities change to do so in virtue of evolving over time. The question this leaves us with is the following: if the world is such that a realization synchronically determines that which it realizes, how, then, should we explain the synchronic realization of free energy minimization? If we go on to accept that the quantities responsible for the production of free energy minimization *change* to minimize free energy, and that these quantities change differently across a timescale of milliseconds, seconds, and minutes, then there can be no such thing as a synchronic (instantaneous) minimization of free energy.

If this is the correct, and if one were to insist on free energy minimization having a realization base, then it seems to me that the best we can do is to say the following: *during that period of time* (however long or short that period of time is), free energy minimization was realized by the quantities specified in figure 11. However, it does not follow from this that during that period of time, the minimization of free energy was synchronically realized by the quantities specifies in figure 11, because synchronicity is not a property that unfolds over time.

Moreover, if the world is such as stated by the free energy principle, then another problem with invoking the realization relation reveals itself by considering, as we have already seen, that a common strategy by which to identify what "constitutes" a realization base of a certain realized property is by appealing to what plays the most salient causal role(s) with regards to the instantiation of the realized property (Cosmelli & Thompson 2010, p. 364). However, free energy minimization is instantiated in nonlinear dynamical systems. According to Cosmelli & Thompson: "In dense nonlinear systems in which all state variable interact with each other, any change in an individual variable becomes inseparable from the state of the rest of the system." (2010, p. 365) Or, as Friston states:

> "[Brain] connections are not static but are changing at the synaptic level all the time. […]. Backward connections are abundant in the brain and are in a position to exert powerful specialization of any area or neuronal population. Modulatory effects imply the post-synaptic response evoked by pre-synaptic input is modulated by, or interacts with, another. *By definition this interaction must depend on non-linear* synaptic and dendritic mechanisms." (2003, p. 1330; italics added)

If we accept that nonlinear dynamics is one fundamental property due to which the patterns of spatiotemporal neuronal assemblies minimize free energy – e.g., by constantly creating predictions about forthcoming sensory events (see e.g., Engel 2010) – it seems a small step to accept that we cannot identify what "constitutes" the realization base of free energy minimization, because the nature of nonlinearity in dynamical systems such as the brain would seem to rule that out (Cosmelli & Thompson 2010, p. 365). One question that this raises is whether the evidence justifies the denial of a physical realization base for free energy minimization. As a first approximation, the answer would seem to be 'no,' in the sense that we would need to have a better understanding of the brain and its nonlinear dynamics to justify the claim that talk of physical realization bases is useless in the context of free energy minimization. However, it does substantiate what I set out to show, namely that we should be cautious with regards to assuming the relevancy of the realization relation in the context of free energy minimization. This leaves it an open empirical question whether there really is a relation of realization between free energy minimization and certain systemic and environmental properties. But what if temporal unfolding or the specific temporal frequencies over which neurons and neuronal assemblies make a significant difference to how the brain gives rise to free energy minimization? That is, suppose free energy minimization is dependent on the integration of multiple different assemblies, and that such integration requires very precise temporal dynamics? Would that makes a difference for how to assess the relationship between realization and free energy minimization? If the realization relation is defined as a synchronic (atemporal) relation between realizer/realized, and if the temporal frequencies over which top-down processing in the brain, say, "match" with bottom-up inputs matters crucially, then this threatens the plausibility of a synchronically conceptualized relation of realization between free energy minimization and the quantities responsible for the production of free energy minimization.

There is ample evidence to suggest that the integration of neuronal assemblies that are involved in top-down processing is dependent upon extremely fast and synchronous activation both in cortico-cortical networks and in cortico-thalamic networks (Engel et al. 2001; Friston & Stephan 2007; Varela et al. 2001; to name a few). For example, Engel et al. (2001) goes as far as to suggest that "top-down factors can lead to states of 'expectancy' or 'anticipation' that can be expressed in the temporal structure of activity patterns before the appearance of stimuli." (2001, p. 710) Other studies suggest that not only changes in discharge rate of neurons or neuronal assemblies, but also changes in neuronal synchrony, can be predictive in nature (see e.g., Riehle et al. 2000). This might be so if the brain utilizes a so-called temporal binding mechanism through which large-scale neuronal assemblies coordinate their activity, as suggested by Engel et al. (2001). If these results hold, it would seem that the world is in fact such that synchronic realization fails in relation to free energy minimization.

Of the different species of realization that I have considered here, wide realization, I have argued, is the most promising view of realization with regards to free energy minimization. However, suppose that the synchronic constraint on realization does fail, then verdict is out on wide realization: it is equally problematic in relation to free energy minimization.

6.6. Conclusion

While it is not possible for me to deny the adequacy of the realization relation for future examinations of the "fit" between realization and free energy minimization, I hope to have made plausible the claim that insofar as the metaphysics of realization remains in the grip of the synchronicity constraint, it is at least prudent to be skeptical about such a potential confluence. Given the conditional character of the argumentational structure in the chapter, the best we can hope for, I think, is that it is unproven that free energy minimization is realized by specific neural and hidden quantities. However, I have argued, if the world turns out to be such as stated in the free energy principle, this does (*prima facie*, at least) give us reasons to doubt the appropriateness of understanding the relation of dependence with regards to free energy minimization as one of realization due to the synchronic conceptualization of the realization relation. To some this will seem an overly negative conclusion. By my lights, however, this is not the case. Rather, it points to future investigations for how to properly ground a clear conceptual framework by which to capture the dynamical ontology of free energy minimization.

7. In search of diachronic ontological emergence

With so much evidence in support of self-organizing, nonlinear, and temporally dynamic processes, crisscrossing levels and boundaries, it is peculiar that many philosophers still hang onto synchronically infused relations of metaphysical dependence. The arguments so far have favored diachronic accounts of dependence relations over synchronic ones. This chapter is no different. In particular, it is now time to examine the heavily debated topic of *emergence*, while casting a critical perspective at accounts of emergence that presuppose that emergence is grounded in the synchronic relation of supervenience (see Broad 1925; Kim 2006; McLaughlin 1997; Rueger 2000; van Cleve 1990; and others, for accounts basing emergence on supervenience).

Given the temporal, nonlinear, and self-organizing nature of dynamical systems and dynamical cognitive systems, what, in principle, stops us from attempting to ground the claim that any robust metaphysics (or, ontology) for such dynamical systems must go exclusively diachronic? In this chapter, what I aim to do is to argue for the following view: in dynamical systems, emergence is best understood as a diachronic ontological relation that does not rely upon synchronic relations such as supervenience.

Insofar as there is to be a third-wave of EC, attempting to provide such an account – of diachronic ontological emergence, where emergence is inconsistent with the supervenience relation – will do a lot of important conceptual work, which will help us move forward into a third-wave, where genuinely dynamical relations are underpinned by a clear conceptual framework of a dynamical ontology. As with the previous chapter, before I set up the arguments to be discussed here, I need to say some things about emergence in general, and its relation to supervenience.

7.1. Sketching emergence

“Emergence” is a philosophical term of art; it can mean, as Kim reminds us, pretty much what you want it to mean, “the only condition being that you had better be reasonably clear about what you mean, and that your concepts turn out to be something interesting and theoretically useful.” (2006, p. 548) Fortunately for us, we do not have to start from scratch when we begin to reflect on how best to understand emergence. For example, there are signposts like the following from Mitchell, who states that: “The key features of emergence for both philosophical treatments and scientific applications are novelty, unpredictability and the causal efficacy of emergent properties or structures, sometimes referred to as downward causation.” (2012, p. 173)

Although Mitchell does not do so, she has a different agenda in her article, for our purposes it is important to note that this passage contains at least three different interpretations of what is at stake in the discussion about emergence. The first meaning is *epistemological*, where emergent properties are usually understood to be unexplainable or unpredictable from knowledge of the lower level system (see e.g., Bedau 1997). The second meaning is *ontological*. In fact, there are two different articulations of ontological emergence in the passage given by Mitchell, even though she does appear to treat them as one and the same in this particular quote. The first is that a system, S, has emergent features, Ps, that are not reducible to the component parts that make up S, or that are not determined solely by the base or lower-level causal processes. As such, the Ps in S may be said to “possess causal capacities not reducible to any of the intrinsic causal capacities of the parts nor to any of the (reducible) relations between the parts.” (Silberstein & McGeever 1999, p. 182)[57] The second meaning is distinct from the first, in the sense that emergent properties may exhibit causal efficaciousness (the first meaning of ontological emergence) and downward causation. As a result, the second sense of ontological emergence is that the Ps have a causal influence on the parts of S that is “consistent

[57] By “intrinsic” I take it that Silberstein & McGeever want to stress that ontologically emergent properties have causal capacities over and above the mere sum of the causal capacities of the individual parts that make up S.

with, but distinct from, the causal capacities of the parts themselves.” (Silberstein & McGeever 1999, p. 182)

The second sense of ontological emergence is the more controversial of the two senses of ontological emergence, yet several philosophers have recently done much to explicate this particular notion by studying downward causation in far-from-thermodynamic dynamical systems (Boogerd et al. 2005; Campbell & Bickhard 2011; Mitchell 2012; Silberstein 2012; Silberstein & McGeever 1999). The distinction between epistemological and ontological emergence may be further taxonomized into either diachronic or synchronic relations of emergence. In this chapter, I only deal with ontological emergence, and will, consequently, leave aside any exposition of epistemological emergence. I will, however, say a bit more about diachronic and synchronic ontological emergence.

Following what O’Connor & Wong (2012) dub “supervenience emergentism,” we can state that the standard view of ontological emergence is cashed out in terms of synchronic supervenience. In his “Emergence: Core ideas and issues,” Kim sets out what he believes to be two necessary conditions for emergence. As Kim says: “The conditions are supervenience and irreducibility.” (2006, p. 548) Or, as Kim puts this formally:

> “*Supervenience*: If property M emerges from properties N_1, …, N_n, then M supervenes on N_1, …, N_n. That is to say, systems that are alike in respect to basal conditions, N_1, …, N_n, must be alike in respect of their emergent properties.” (2006, p. 550; italics in original)

Consider, also, what Rueger says: “Robust supervenience, I argue, […] provides a natural background for reconstructing the notion of (diachronic) property emergence in a way acceptable to physicalists.” (2000, p. 466) Or, as McLaughlin states: “If *P* is a property of *w*, then *P* is emergent if and only if (1) *P* supervenes with nomological necessity, but not with logical necessity on properties that parts of *w* have taken separately or in other combinations […].” (1997, p. 39) There are differences between Kim and McLaughlin, on the one hand, and Rueger, on the other. However, what matters for my purposes is that all three authors explicate emergence as a supervenience relation.

In contrast to synchronic supervenience emergentism, whose proponents usually argue that given the subvenient or basal conditions at a particular time instant *t*, there will be some emergent property *M* at *t*, defenders of diachronic ontological emergence stress that (i) the historical or temporal trajectory of the system matters ineliminably in determining which emergent properties are instantiated in a system over time, and (ii) that grounding emergent properties by appeal to the relation of supervenience is wrong (Bickhard 2004; Campbell & Bickhard 2011; Kirchhoff 2013c; O'Connor 2000; Seibt 2009). As a result, for the advocate of diachronic ontological emergence, there will be an emergent property *M* in a system *S* given (i) certain dynamical lower-level processes of S that unfold over time, and (ii) in virtue of the fact that *prior systemic states and processes* play an ineliminable role in determining which emergent properties come about over time and at a time.

7.2. Arguments

My aim in this chapter is to argue the following: in dynamical systems, emergence is best understood as a diachronic ontological relation that does not rely upon synchronic relations such as supervenience. Thus, I join the ranks of those that (i) favor diachronic ontological emergence, and (ii) are suspicious of the standard view of emergence as a supervenience relation. With regards to the two different senses of ontological emergence outlined above, the kind of ontological emergence that I shall discuss and defend is ontological emergence with downward causation. In doing so, I will argue for two points.

First, claims involving emergence are ubiquitous in discussions of self-organizing, nonlinear dynamics (Beer 1995; Chemero 2009; Kelso 1995; van Gelder 1998), artificial life and robotics research (Brooks 1999; Pfeifer et al. 2005), extended and distributed cognition (Clark 1997; Hutchins 1995; Menary 2007; Wheeler 2005), enactivism (Di Paolo 2009; Varela et al. 1991), developmental systems theoretic approaches in biology (Griffiths & Stotz 2000; Oyama et al. 2001), and certain divisions in philosophy of science (Boogerd et al. 2005; Wimsatt 1986). As a result, my first argument is that what these sciences and theories show us is that supervenience emergentism cannot explain *how* higher-level emergent phenomena arise and are *maintained* over time, because the supervenience relation is unable to analyze the context-sensitive, nonlinear, and temporal dynamics that are characteristic of dynamical systems and their emergent properties.

Second, if the world is such that emergent properties in dynamical systems do not coincide with the supervenience relation, then it seems to me that there is a potential tension within certain accounts of EC. That is, some defenders of EC use both ontologically diachronic construals of emergence and the supervenience relation to argue for EC. However, many (though not all) defenders of EC take extended cognitive systems to be dynamical systems. But, suppose that it is correct to state that the emergence of cognitive properties in dynamical systems (when those dynamical systems are understood as extended cognitive dynamical systems) does not coincide with relations of supervenience, then it follows, I submit, that the appeal to the supervenience relation is misplaced. Consider, for example, that in his many writings across dynamical, embodied, and extended cognition, Clark (1997, 2001, 2008) has done much to develop the idea that the study of diachronic ontological emergence must sit at the center of the scientific pursuit to understand cognition. However, Clark also defends claims of the following variety:

> "A recurrent theme in previous chapters has been the ability of body and world to act as what might now be dubbed "participant machinery" – that is, to form part of the very machinery by means of which mind and cognition are physically realized and hence form part of the local material *supervenience base* for various mental states and processes." (Clark 2008, p. 207; italics added)

Perhaps Clark is justified in asserting that neural, bodily, and worldly elements are all part of the local material supervenience base with regards to various mental *states*, since "states" in the philosophy of mind has commonly been understood to express a *discrete* mental state instantiated or exemplified at a particular time *t* (see e.g., Kim 2011). Suppose that this is the correct way to articulate the relationship between certain physical states and a mental state, then the relation of supervenience might be appropriate, since the latter indicates that this relationship – between a mental state and the physical states realizing that mental state – is synchronic. By extension, if it turns out that mental states emerge from certain physical states, and if the relation of supervenience adequately expresses the relationship between this mental state and the physical states, then it follows that we have a case of supervenience emergentism.

However, it is not clear that the claim that *processes* synchronically supervene on a local material subvenient base is justified, because whereas supervenience is a synchronic relation, processes are – I have argued throughout the thesis – inherently diachronic, *viz*., in order to persist, processes require continuity in space and time (Goldie 2011; Hofweber & Velleman 2011; Kirchhoff 2013a; Seibt 2009; Spivey 2007). If it turns out to be correct that emergent cognitive processes do not coincide with supervenience, then emergent cognitive processes do not supervene on an extended subvenient base distributed across parts of the brain, body, and world. I do not address whether mental states supervene on a subvenient physical base in this chapter. Rather, in terms of the second argument to be developed here, I restrict my attention to the problem of positing emergent cognitive *processes*, on the one hand, and claiming that such processes supervene on certain lower-level physical elements, on the other.

7.3. Overview

In section 7.4, I consider a couple of examples of emergence in dynamical systems to serve as a backdrop for the discussion to come. In section 7.5, I analyze the relation between emergence and supervenience. While I will spend most of my energy on discussing the philosophical arguments provided by Kim (1999, 2006), whose views on this relationship have been influential in shaping the debate about emergence, I also consider Rueger's (2000) argument that diachronic emergence in dynamical systems coincides with the supervenience relation. During this discussion, I use the test cases as a basis for arguing the following. First, Kim is wrong to insist that supervenience is necessary for emergence. Second, Kim is wrong in his rebuttal of emergence, because downward causation is less problematic than Kim thinks. And third, that Rueger's argument for why diachronic emergence coincides with supervenience ignores that the dependence relation between higher- and lower-level activities in dynamical systems may be *diagonal* rather than strictly vertical as presupposed by the supervenience relation. It is my hope that these three discussions will be enough to establish the justifiability of diachronic ontological emergence. Finally, I discuss the second argument outlined above, namely that there is a potential tension in the literature on EC given the fact that certain defenders of EC use both diachronic ontological emergence and supervenience to argue for EC.

7.4. An example

An understanding of emergence that has become widespread in the scientific and philosophical models listed above – nonlinear dynamics, artificial life and robotics research, extended/distributed cognition, enactivism, developmental systems theory in biology, and certain strands of philosophy of science – is that emergence is identified with a certain kind of non-aggregative relationship, displaying both non-linearity and self-organization. That is, a phenomenon will become increasingly less aggregative and consequently more likely emergent as the number, duration, and complexity of the interactions between its parts increases (Wheeler 2005, p. 260; Wimsatt 2000, p. 275) Aggregation we can understand as a simple kind of compositional relationship between a whole and its parts, e.g., the weight of a pile of rocks is the aggregate weight of each component rock making up the whole. However, as Mitchell points out (2012, p. 179), there are different ways in which compositional aggregation will fail to pick out the appropriate relationship between a whole and its parts.

One way is found in the complexity of dynamical systems exemplified by self-organization and nonlinearity. Consider the example of a fluid heated from below and cooled from above. Take some oil, put it in a pan, and apply a heat source from below. As the heat is applied it increases the difference in the temperature between the top and the bottom of the oil layer. At a critical threshold, an event called "an *instability*" occurs such that the liquid begins to self-organize a coherently rolling motion. This motion is a *convection roll*. What happens is that the cooler liquid at the top is denser, thereby falling, whereas the liquid at the bottom is warmer (and so lighter), thus tending to rise to the top. Of such a process, Kelso says:

> "The resulting convection rolls are what physicists call a *collective* or *cooperative* effect, which arises without any external instructions. The temperature gradient is called a *control parameter* [but it does not] prescribe or contain the code for the emerging pattern. […] Such spontaneous pattern formation is exactly what we mean by *self-organization*: the system organizes

> itself, but there is no "self," no agent inside the system doing the organizing. [And] the amplitude of the convection rolls plays the role of an *order parameter* or *collective variable*: all the parts of the liquid no longer behave independently but are sucked into an ordered coordinated pattern […]. It is this coherent pattern that is described by the order parameter and it is the order parameter dynamics that characterizes how patterns form and evolve over time." (1995, pp. 7-8; italics in original)

How is "emergence" used here? Is the emergent property an object, a mechanism, a thing or a process? In the case of convection rolls, the emergent phenomenon is a process, in the sense that it is an unfolding dynamical pattern. Is the emergent pattern novel? With regards to emergent properties, novelty is typically explained as either the result of unpredictability or unexpectedness, or both. However, convection rolls are not novel, because they are unexpected. The reason for this is that convection rolls are not unexpected: given similar start-up conditions, only a small variation in the value of the control parameter (the temperature gradient) leads to qualitatively similar behavior across different token phenomena. Instead convection rolls are novel in virtue of being unpredictable. To get a grip on the idea of "unpredictability," consider how Kim (1999) distinguishes (importantly) between *inductive* predictability and *theoretical* predictability. Convection rolls are inductively predictable, in the sense that given certain start-up conditions, varying the control parameter will typically result in the emergence of convection rolls. However, what defenders of emergence are denying is the theoretical predictability of convection rolls given our knowledge of the 'basal' constituents and the nonlinear and self-organization of constituents over time. The emergence of convection rolls is novel, because they are theoretically unpredictable.

In what sense is the emergence of convection rolls ontological? Or, is it merely epistemic? Consider, first, what Silberstein & McGeever say about epistemological emergence:

> "A property of an object or system is epistemologically emergent if the property is *reducible* to or *determined* by the intrinsic properties of the ultimate constituents of the object or system, while at the same time it is very difficult to explain, predict or derive the property on the basis of the ultimate

> constituents. Epistemologically emergent properties are novel only at the level of description.” (1999, p. 186)

The fact that the appearance of convection rolls is novel due to its theoretical unpredictability would suggest that it is emergent epistemologically, and only novel at the level of description. Obviously, if this is the case, then this example cannot be a case of ontological emergence, in the sense that the definition of epistemological emergence precludes ontological emergence. Recall that ontological emergence is the case when emergent phenomena are not reducible to their ultimate constituents, and may either exhibit causal efficaciousness or downward causation, or both. This example, in addition to theoretical unpredictability, exhibits downward causation. As Kelso states: “the amplitude of the convection rolls plays the role of [a collective variable]: all the parts of the liquid no longer behave independently but are sucked into an ordered coordinated pattern […].” (1995, p. 8) That is, the convection cycle involves a kind of “circular causation” (Kelso 1995, p. 9; see also Clark 2001) in which the self-organized dynamics of the individual component molecules give rise to a larger pattern, and this pattern, in turn, enslaves the individual molecules into a recurrent pattern of rising and falling. Therefore, we get both epistemic emergence and ontological emergence.

Can we be more specific than this? We can ask whether the kind of emergence is diachronic or synchronic, or both? Since this question is the central point of discussion of this chapter, and because I will provide a lengthy discussion of this question in the following sections, I will only motivate a few points here. Consider that while the supervenience relation specifies a relation between certain properties of a system *at a particular time instant t*, convection rolls emerge in a system in virtue of a *change* in the qualitative behavior of a dynamical system *over time*. In dynamical systems theoretical terms, there is a qualitative change in the system’s phase space portrait due to differences in the way in which the components interact and a change over time.

Suppose, as Rueger does, that the phase space portrait remains what he refers to as “qualitatively the same under perturbations of the dynamics, i.e., small variations in the value of p [the control parameter], [then] the system is structurally stable. If the perturbations generates a qualitatively different portrait of trajectories, the system is structurally unstable.” (2000, pp. 472-73) The example of convection rolls is structurally unstable, in the sense that a relatively small variation in the temperature gradient results in a qualitatively different phase space portrait, *viz*., from a phase space with no convection rolls to a phase space with convection rolls. Furthermore, for Rueger, insofar as the system is structurally unstable, the emergent properties, in that system, are diachronically emergent. But, while Rueger goes on to argue that even diachronic emergence is a supervenience relation, in what follows, I consider a number of different ways by which to criticize supervenience emergentism.

I start this discussion by providing a detailed examination of Kim's (1999, 2006) arguments intended to show that (a) supervenience is a necessary (but not a sufficient) condition for emergence, and (b) that emergence is ultimately problematic due to problems with downward causation. Having done this, I discuss the account given by Rueger for why diachronic emergence is a supervenience relation. Before I can discuss Kim's arguments, we need to know what they are. I begin by exposing Kim's argument that the supervenience relation is necessary for emergence (section 7.5), followed by a discussion of this argument (section 7.5.1). Then I expose Kim's argument against downward causation (section 7.6), followed by a discussion of this argument (section 7.6.1). Similarly with Rueger, where I start with an exposition of Rueger's argument (section 7.7), followed by a discussion of this argument (section 7.7.1).

7.5. Kim's argument that supervenience is necessary for emergence

Kim argues that supervenience is a necessary condition that any adequate analysis of emergence must satisfy. Consider two wholes that have identical microstructure (i.e., they are composed of identical basic physical constituents configured in an identical structure) yet different concerning the supposedly emergent properties. Applying this to the dynamical case of rising and falling fluid, there may be two molecule-for-molecule identical systems, although only one of them displays convection rolls. If such cases were possible, Kim says, the relation between putatively higher-level emergent phenomena and their physical base components would be "irregular, haphazard [and/or] coincidental." (2006, p. 550) In other words, what justification could there be for saying that cyclical convection rolls emerge from that specific physical configuration rather than another? Hence, Kim rejects the view that an emergent feature of a whole is not *determined* exclusively by bits of matter, in a certain structural organization. As Kim says: "If supervenience, or upward necessitation, is taken away, that takes away something essential to the meaning of "emergence" […]." (2006, p. 550) Therefore, he concludes by stating supervenience is a necessary condition for emergence; that we have to accept the following:

> "*Supervenience*: If property *M* emerges from properties $N_1, \ldots, N_n$, then *M* supervenes on $N_1, \ldots, N_n$. That is to say, systems that are alike in respect of basal conditions, $N_1, \ldots, N_n$ must be alike in respect of their emergent properties." (2006, p. 550; italics in original)

Kim argues that if *M* emerges from $N_1, \ldots, N_n$, those properties (*M* and its emergent base properties) must meet the supervenience condition. Dovetailing his account of supervenience, with the principle of synchronic determination, Kim advances this additional specification: "Supervenience/determination: Property *M* supervenes on, or is determined by, properties $N_1, \ldots, N_n$ in the sense that whenever anything has $N_1, \ldots, N_n$, it necessarily has *M*." (2006, p. 550)

7.5.1. Discussion of Kim's "supervenience is necessary for emergence claim"

With this description of Kim's account of the relationship between emergence and supervenience, we are now in a position to critically discuss the question whether emergent features necessarily supervene on their physical features.

Does the emergence of convection rolls necessarily supervene on its base-level parts and their arrangement? To get us started, I present an alternative account of emergence that takes emergence to be ontologically diachronic and non-supervenient, and then apply our test case from dynamical systems to show that emergent phenomena do not necessarily supervene on their lower-level physical components and their configuration. The alternative account of emergence I wish to consider here is due to O'Connor (2000) and O'Connor & Wong (2005), who argue that the standard view of emergence as a supervenience relation is suspect (O'Connor & Wong 2012, p. 14). In O'Connor (2000), the target is the relation between a neural mechanism and the emergent properties that this neural mechanism gives rise to. Here I substitute O'Connor's example with my discussion of convection rolls. This is the argument provided by O'Connor:

Suppose that when a dynamical mechanism *H* comes to have a particular configuration *P** at time *T0*, the baseline emergent feature *E* is the result at *T1*. Because *H* is a dynamical system, the specific organization of *H* is, in part, at least, determined by *P** as it unfolds. I say that *H* is " in part, at least" determined by *P** during the period of *H's* existence. That is because *H* qua being a dynamical system is an *open*, *non-equilibrium system*. H is open, in the sense that it interacts with its environment – exchanging energy and matter. H is non-equilibrium, in the sense that without these interactions, it cannot maintain proper (low) levels of physical entropy (Kelso 1995, p. 4). Thus, we cannot presuppose that *H* is fully determined by *P** at any point in time, e.g., at *T1*. *P0* is the remaining aspect of *H's* internal states at *T0*, and *P@* is the summation of factors in *H's* immediate environment that have an influence on the state of *H* at *T1*. Let "→" represent the causal relation between component properties. This gives us the following relationships (O'Connor 2000, p. 111):

- *P** at *T0* → *E* at *T1*

  And
- *P* + P0 + P@* at *T0* → *P* + P1* at *T1*

This last conjunction highlights the total internal state of *H* at *T1*, where *P1* is the remainder beyond *P**. In the language of dynamical systems theory, we can understand the total internal state of *H* to be sets of *interdependent variables*, where a variable is a simple entity that can have a different state or value at different times (van Gelder 1998, p. 616). According to O'Connor, *E* at *T1* will be a joint determiner of *H's* physical configuration at *T2*, but not *H's* continuing to have *P**. Moreover, *E* may also help in determining, at *T2*, the emergence of another feature, *E2*. On O'Connor's account, this looks diagrammatically as follows (Fig. 12):

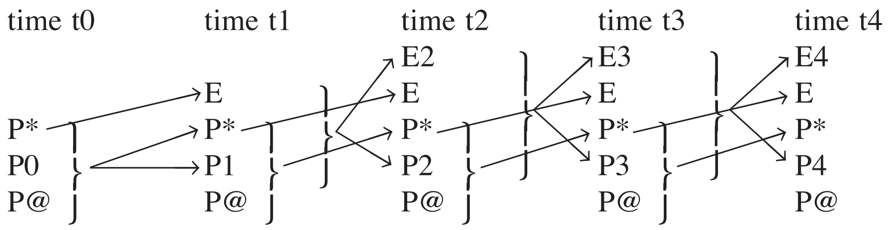

Fig. 12: Diagram of the evolution of emergent features (adapted from O’Connor 2000, p. 111)

Kim’s invocation of supervenience to deliver a non-coincidental conception of emergence raises the question whether there can exist two systems that are alike in respect of lower-level features, yet that are not alike in respect to their emergent features. Indeed, if such a possibility turns out to exist, which O’Connor’s formal argument claims that it does, this suggests the failure of supervenience in accounts of ontologically diachronic emergence.

In our test case, *H* is a dynamical system comprised of *P**, *P@* (in this case, *P@* is the source of heat applied), and *P0* at time *T0*. Because of the inherent nonlinearity of *H*, a liquid heated only weakly from below will not display any macro-motion. By “nonlinearity” I mean that if there is an interaction between two components, X and Y, say, the value of Y does not increase proportionally to the value of X, but may remain at zero, e.g., until X reaches a certain critical threshold (Clark 2001, p. 115). However, at a critical value of the temperature gradient, the liquid will start to show the familiar macroscopic rolling motion we know as convection rolls. In the language of dynamical systems, here the liquid is in a bi-stable dynamical state such that the convection rolls may rotate in one direction or in the other direction. That is, either from right to left or from left to right in its rolling motion. The direction of the rolling motion that emerges is random[58]. This is due to the onset of a *bifurcation*. Before the onset of the rolling motion, when the overall value of *H* is below the critical order parameter value, *H* can only be in one possible state and that is a state of rest – diagrammatically shown by the liquid being in a stable (deep) energy well. At the instability point, a bifurcation (or

[58] Microscopically, no two examples will be exactly alike. It is the “motion” in this example that reflects the physical entropy production.

branching) occurs such that *H* moves away from its stable state to a bi-stable state – diagrammatically shown by the potential of the liquid to be in one of two dynamical states. That is, "two rolling motions whose rotation speed is equal but opposite emerge spontaneously, [with only] one, of course, realized in the experiment." (Kelso 1995, p. 10) What was previously a stable state becomes an unstable state.

Consider the status of the baseline emergent feature *E*, with respect to the times *T0* and *T1* in O'Connor's diagram. Here *E* is absent at *T0* but present at *T1*. From this specification, O'Connor points out the following: "The underlying physical [features] are different, too, but that is not the reason for the difference in emergent [features]. For the differentiating factors (*P0*, *P1* and […] *P@*) are, by hypothesis, not directly relevant to the occurrence of *E*. *P** alone is so relevant." (2000, p. 112) It might sound strange to omit *P0*, *P1*, and *P@* from consideration in this discussion. Why does O'Connor do this? On the one hand, Kim clearly states the idea of emergence is the thought that a physical system, composed of only bits of matter and their organization, can begin to exhibit emergent features. This is what *P** implies. On the other hand, *P0* and *PI* represent any remaining physical elements that do not participate in the generation of emergent phenomena, where *P@* is a parameter outside of *H*. For readers familiar with dynamical systems theory, the omission of certain parameters, because they are not directly relevant to the occurrence of the phenomenon in question, should be familiar (see e.g., Beer 1995, p. 181).

With this clarification out of the way, note that *E* is absent in *T0*, but present immediately thereafter in *T1*. Consequently, in the first instantiation of *H*, no emergent features will be present. Even if *H* at *T0* has reached the point of instability, this process has to unfold over a period of time; hence, even if *H* begins the branching movement at *T0*, yet sharing with *H* at *T1* what is specified in *P**, only one of the physical configurations has *E* (O'Connor 2000, p. 112). In other words, let us say we have two dynamical systems both of which have alike physical configuration, represented by *P**. In one system the interaction of the molecules is faster, because of the higher temperature gradient, where this is not the case in the second system; in one system there is rolling motion and in the other system there is not rolling motion; but the arrangement of material bits might be the same across these two systems at a specific instant.

Rolling motions are not the only possibilities. In an open container, for instance, surface tension can also affect the flow. According to Kelso: "The net effect of this force [surface tension] is to minimize the surface area of the fluid, causing tessellation of the surface and the formation of hexagon cells. In the center of each hexagon, liquid rises, spreads out over the surface, and sinks at the perimeter where the hexagons join." (1995, pp. 7-8) Plausibly, a couple of dynamical systems might exist that are alike with respect to *P** at *T0*, but where one has *E* (the cyclic pattern of convection rolls) at *T1*, whereas the other has *E** (the pattern of liquid rising, spreading out over the surface, and sinks at the perimeter) at *T1*. This is plausible, and can be explained by taking into account the mechanisms involved in the generations of these two different emergent patterns. As Kelso explains: "An important point is that two quite distinct mechanisms – one to do with buoyancy and the other surface tension – can give rise to [either] the same pattern [or to two different patterns]." (1995, p. 8) Crucially, neither buoyancy nor surface tension are elements of *P**, since *P**, even on the account provided by Kim, is a physical system composed only of bits of matter and their organization; however, buoyancy and surface tension is neither bits of matter nor the organization of bits of matter – both mechanisms are the *result* of a system organized and unfolding in a particular manner.

If this is true, O'Connor is correct to insist that supervenience emergentism is problematic. Once we consider the contribution of a system's *prior* trajectory through its dynamical state space, we get the picture that prior systemic elements play a role in determining emergent features of a system, and that emergent features themselves are temporally extended in time. Unfortunately for those insisting on the compatibility of emergence and supervenience, the relation of supervenience fails to apply to cases in which the relata of emergence are ontologically diachronic.

Consider, furthermore, that the supervenience/determination condition states that if some feature *E* of a system S supervenes on, or is determined by, features N_1, …, N_n, then *S* necessarily has *E*. However, what the example of convection rolls shows is that there can be a divergence between the emergent features of two dynamical systems in the face of physical similarity at the base level. As O'Connor puts it: "You might have the underlying physical properties *P** and *P2* without having had *E2*. For *E2* is a product of the immediately prior state of *H* (comprised of *P**, *P1*, and *E*). This prior state presumably could have been different (such that

E2 would not occur at *t2*) […].” (2000, p. 112) Because dynamical systems are open and undergo bifurcations over time, open systems may instance the same physical properties here-and-now, while instantiating different emergent properties (Kelso 1995; O’Connor & Wong 2012).

The example of convection rolls is not special, in the sense that its main implications for our understanding of emergence are non-generalizable to other phenomena. At this stage, and to provide additional reasons for my skepticism against emergence as a supervenience relation, I consider a different example from which I draw out two problems with the assumption that emergence is a supervenience relation.

Consider what evidence regarding the molecular structure of water indicates about the part-whole relation between the macroscopic pattern of water and its component parts, here molecules of water. If the macroscopic pattern of water emerges from its lower-level physical constituents, and if one further assumes that this relation of emergence is supervenient, this entails that the macroscopic pattern of water has its features *synchronically*. However, water, Ladyman & Ross inform us “is composed of oxygen and hydrogen in various polymeric forms, such as $(H_2O)_2$, $(H_2O)_3$), and so on, that are *constantly forming, dissipating,* and *reforming over short time periods* in such a way as to give rise to the familiar properties of the macroscopic kind water.” (2007, p. 21; italics added) That is, water, as Ross & Ladyman baldly state, is an emergent feature of a complex dynamical system, and it “makes no sense to imagine it having its familiar properties *synchronically*.” (2010, p. 160; italics added)

Insofar as water has emergent features, the presupposition of the standard and representative view of emergence – that this is a supervenience relation – fails to capture the dynamics of how the higher level *is* emergent and how it is maintained as it unfolds over time. Furthermore, considering the different phases of water, and related thermodynamic considerations of how the macroscopic whole “water” relates to its parts, van Brakel begins by asking whether water consists of H_2O? At a first approximation, water is not composed of H_2O, since the part-whole composition of water is *context sensitive* and *contingent*. By “context sensitive and contingent,” I follow Oyama et al. (2001) and take this to mean that the “significance of any cause is contingent upon the state of the rest of the system.” (2001, p. 2). For instance, in liquid water, water molecules form larger clusters via interaction. Or, in ice, as van Brakel points out, "there aren't really individual molecules. [In fact, in] solid neon one might [even] come across hexamers." (2010, p. 131)[59] But if the different phases of water reflect differences in their microstructure, one might think that surely it is possible to identify the relevant molecules as being *like this*? However, according to van Brakel, water "contains isotopic variants such as HDO and D_2O. The physico-chemical properties of deuterium oxide are different from those of ordinary water, and therefore, its role in biochemical processes." (2010, pp. 131-32) Hence, the answer is that we cannot identify water molecules as being "like this". But what if we restrict our discussion to only one set of isotopes, then surely all H-atoms (hydrogen atoms) in water are identical? Again van Brakel answers this question with a straightforward "no," since there "is [both] ortho- and para-hydrogen, which have different physico-chemical properties." (2010, p. 132) All right the reader might wonder; nevertheless, even if the part-whole relation between water and its constituents is context sensitive, is it not still the case that a water molecule always *behaves* as a H_2O molecule? Fascinatingly, according to van Brakel the answer to this is: "No! […] in some circumstances, water molecules are not "seen" as H_2O by entities with which it interacts, but more like something like H_1, $_5O$." (2010, p. 132)[60] We cannot even assume that water always consists of interactions between H and O atoms. Indeed, "there is some probability (however, small) that a water molecule will suddenly transform into a Neon atom." (Belyaev et al. 2001; quoted in van Brakel 2010, p. 132)

Thus, one problem with the standard account of emergence as a supervenience relation is that because of its commitment to supervenience, and since this metaphysical relation is a synchronic relation, it misses out on the dynamics by which oxygen and hydrogen, in various forms such as $(H_2O)_2)$, $(H_2O)_3$, and so on, are constantly forming, dissipating, and reforming over very short time periods,

[59] A hexamer is considered to be the smallest drop of water, because it is the smallest cluster of water that is three-dimensional.

[60] "Schewe et al. (2003): A water molecule's chemical formula is really not H_2O, at least not from the perspective of neutrons and electrons interacting with the molecule for only attoseconds (less than 10^{-15} s)." (Cited in van Brakel 2010, p. 132; footnote 42)

thereby giving rise to the so familiar features of the macroscopic kind water (Ross & Ladyman 2010, p. 21).

This points us in the direction of a related problem, which comes about through recognizing that as a part-whole relation, the standard account of emergence as a supervenience relation presupposes that relational (i.e., extrinsic) properties supervene on non-relational (i.e., intrinsic) properties of its lower-level, subvenient base (Lewis 1983; McLaughlin & Bennett 2011). "Intrinsic properties" are normally understood to be those properties that an object may possess independently of everything else that exists, or independently of whether or not anything else exists. For example, the charge and mass of a classical particle are thought to be intrinsic properties of an object, whereas "being south of Sydney" is dependent on someone's relation to something else. In the case of water, supervenience fails to obtain. It fails because of the context sensitivity and contingency of the part-whole relations characteristic of water. If emergent wholes come about through continuous and reciprocal interaction between components at the emergent base level, and if the features of the parts cannot be determined independently of the context in which they are embedded, then this challenges the idea that relational features supervene on non-relational features. It challenges the very foundation presupposed by supervenience relations, namely that extrinsic stuff supervenes on intrinsic stuff (Humphreys 1997; Ladyman & Ross 2007; O'Connor 2000).

7.6. Kim's argument against downward causation

If it turns out to be correct that at least in cases of dynamical systems, emergence does not by necessity presuppose supervenience, Kim has a second argument that threatens to render the notion of emergence incoherent. As Kim rightly notes, concerning ontological emergence, it "is critically important to the emergentist that emergent properties have distinctive *causal powers* of their own, *irreducible* to the causal powers of their base properties." (2006, p. 557; italics added) Otherwise, it would seem, that putative emergent phenomena would be mere epiphenomena. Any account of ontological emergence, including mine, must thus be committed to either emergent properties that are causally efficacious or to downward causation, or to both.

Across numerous publications, Kim has repeatedly used the following argument to establish that all forms of non-reductive physicalism (e.g., emergence), which separate higher-level features and their compositional lower-level components and activities collapses to some form of reductive physicalism. If this argument is successful, it leaves no room for ontological accounts of emergence.

Here is my iteration of Kim's argument: Suppose there is an emergent feature, *M*, with causal powers, and that some instantiation of it brings about another emergent feature *M**. But, *ex hypothesi*, *M**, as an emergent feature, must be composed by its physical base *P**, without which *M** would not be present. The initial point that Kim wants to make is that for this story to be coherent, we must accept that the instance of *M* caused *M** to be instantiated by causing the instantiation of *P**. However, as an emergent, *M* must itself have a physical base, call this base *P*. Furthermore, if *M* supervenes on *P*, then the instantiation of *P* should be sufficient for the instantiation of *M*. At this juncture in the argument, Kim invokes the principle of transitivity to assert the following: if *M* is causally sufficient for *P**, and thereby *M**, then *P* is causally sufficient for both *P** and *M**. Hence, as Kim points out: "This appears to make the emergent property *M* otiose and dispensable as a cause of *P**; it seems that we can explain the occurrence of *P** simply in terms of *P*, without invoking *M* at all. […]. If *M* is somehow retained as a cause, we are faced with the highly implausible consequence that every case of downward causation involves causal overdetermination (since *P* remains a cause of *P** as well)." (2006, p. 558)

Kim concludes his argument against emergence by stating that if this argument cannot be successfully rebutted, it threatens to bankrupt the central idea of ontological emergence, namely that emergent features can have causal efficacy over and above those of the components making up the base level. As he says: "If downward causation goes, so goes emergentism." (2006, p. 558)

7.6.1. Discussion of Kim's argument against downward causation

The self-organization and nonlinearity that is characteristic of the dynamical complexity in dynamical systems, like the case of convection rolls and water, reveals that while deterministic, the emergent behaviors are unpredictable given their sensitivity to variations in starting and evolving boundary conditions (Mitchell

2012, p. 181). But this gives us only epistemological emergence, whereas Kim's argument against emergence is leveled at ontological emergence with downward causation. To problematize Kim's critical argument, I need to show that the idea of downward causation is not as problematic as Kim thinks. Indeed, as Campbell & Bickhard (2011) point out, there is no shortage of examples of downward causation. Consider, e.g., Couzin & Krause's (2003) identification of emergence in flocking starlings:

> "[It] is usually not possible to *predict* how the interactions among a large number of components within a system result in population-level properties. Such systems often exhibit a recursive, nonlinear relationship between the individual behavior and collective ('higher-order') properties generated by these interactions; *the individual interactions create a larger-scale structure, which influences the behavior of individuals, which changes the higher-order structure*, and so on." (2003, p. 3; italics added)

Harking back to the language of dynamical systems, an order parameter (or collective variable) comes about through the coordination between individuals (birds, molecules, etc.), which have an influence on the parts. What the contemporary scientific sense of emergence gives us are concrete accounts of how and why multiple top-down and bottom-up, forward and reverse feedback, loops (a) can result in the theoretical unpredictability of emergent phenomena, and (b) the existence of wholly natural modes of downward causation. As Clark says about the continuous self-organizing, top-down and bottom-up loops in dynamical systems: "This cycle involves as kind of "circular causation" in which the activity of the simple components leads to a larger pattern, which then *enslaves* those same components, locking them into a cycle [in the case of convection rolls] of rising and falling." (2001, p. 112)

While I am sure some will find the validity of Kim's logic persuasive, there are assumptions made that are open to criticism. First, Campbell & Bickhard argue that Kim's argument begs the question. That is, it says that *P* is nomologically sufficient for M. However, since this is precisely the claim of physicalism, it cannot be used as a premise in an argument against the possibility of ontological emergence. Even if this begs the question, there is a related problem, namely: even if the conditions arising at a higher level are constituted by a particular configuration of components and their activities, and if the latter is nomologically sufficient for the emergence of the former, this does not rule out the possibility that once generated, the higher-order features may "act back" upon, and thus shape, the features of the base-level constituents. Second, Kim's argument against downward causation is built on the justifiability of supervenience. But, as we have seen, this premise is problematic. As O'Connor stresses: "there might be two objects having identical […] physical properties […] and existing in the same external circumstances, yet one [is emergent] and the other [is not]." (2000, p. 112)

7.7. Rueger's argument that diachronic emergence coincides with supervenience

In a paper pertinent to my claim that diachronic ontological emergence, in particular when diachronic and ontological properties emerge in dynamical systems, does not rely on the supervenience relation, Rueger (2000) defends the opposite position, that even in dynamical systems, diachronic and ontological emergent properties are supervening on a subvenient base. Rueger provides two interrelated arguments for this claim.

Take a physical system here-and-now, and describe it (in the terms of dynamical systems theory) by way of some set of *generalized coordinates* (that is, position and momentum). These generalized coordinates take on a sequence of values over a time interval. This sequence of values is represented as a *trajectory* (i.e., a set of positions in the state space through which the system might pass successively) in the system's *phase space portrait* (i.e., the representation of the system). Alongside these variables, *control parameters* (i.e., a parameter of a system whose continuous quantitative change may lead to noncontinuous, qualitative change in the phase space portrait of the system) such as a heat gradient (in our case of convection rolls) specify features of the system that are not determined by the system's internal dynamics. As Rueger says: "The combination of generalized coordinates and control parameters forms the *subvenient base* of the system. The phase space trajectories express the behavior of the system – the collection of its *supervenient properties*." (2000, p. 472) For simplicity, consider the following illustration provided by Rueger:

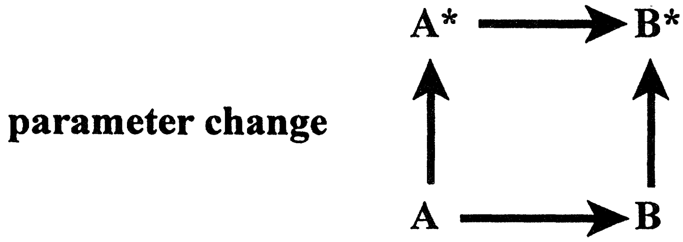

Fig. 13 Temporal evolution of a system {A, B} into {A*, B*} (adapted from Rueger 2000, p. 473)

The vertical arrow represents the synchronic supervenience relation between some properties of a system *at an instant t*, whereas the horizontal arrow represents qualitative change over a time interval, e.g., from t_1 to t_2. The first argument Rueger gives for supervenience emergentism turns on *structural stability*. A physical system is structurally stable: if a set of A*-properties supervenes on a set of A-properties, then "wiggling" some member(s) of the set of A-properties will not result in a qualitatively different set of B*-properties, which given the temporal evolution of the system supervene on a set of B-properties (Rueger 2000, p. 476). That is, if A*-properties supervene on A-properties at t_1, then a small change in A-properties – e.g., a small change in the value of the control parameter – will not result in a qualitatively different phase portrait at t_2 such as illustrated in the relation between {B, B*}. In other words, if A*-properties supervene on A-properties, the continuous change of a control parameter will not lead to a noncontinuous, qualitative change in the phase space portrait of the system. If this is the case, then the set of A*-properties supervene on the subvenient A-properties at t_1, and despite small changes in A-properties, this supervenience relation will remain as it is at each point in time over an interval.

Even if physical systems may be structurally stable, instabilities do occur, and the *robustness requirement* of the first argument is violated. That is, the condition of structural stability is violated. If this is the case, then, according to Rueger, the overall evolution of the system exhibits the emergence of *diachronically* emergent properties. The second argument that Rueger provides is set up to establish that even though the condition of structural stability is violated, and, as a result, the system exhibits diachronically emergent properties, structurally unstable systems are nevertheless explicable in terms of the supervenience relation. Provisionally we need to distinguish between *individual systems*, in the sense that *structural stability* (on Rueger's account) is defined for individual systems (2000, p. 477) and *families of systems* intended to show that even though a given *token* system may be structurally unstable (under some perturbation), the system *S* could nevertheless belong to a stable *type* or family of systems. According to Rueger: "If it does, then the bifurcation or instability in the system occurs in a structurally stable way: all the systems in the family exhibit qualitatively the same bifurcation behavior." (2000, p. 479)

7.7.1. Discussion of Rueger's two arguments

To discuss Rueger's two arguments for why diachronic emergence coincides with supervenience, let us consider our familiar example of convection rolls (figure 14):

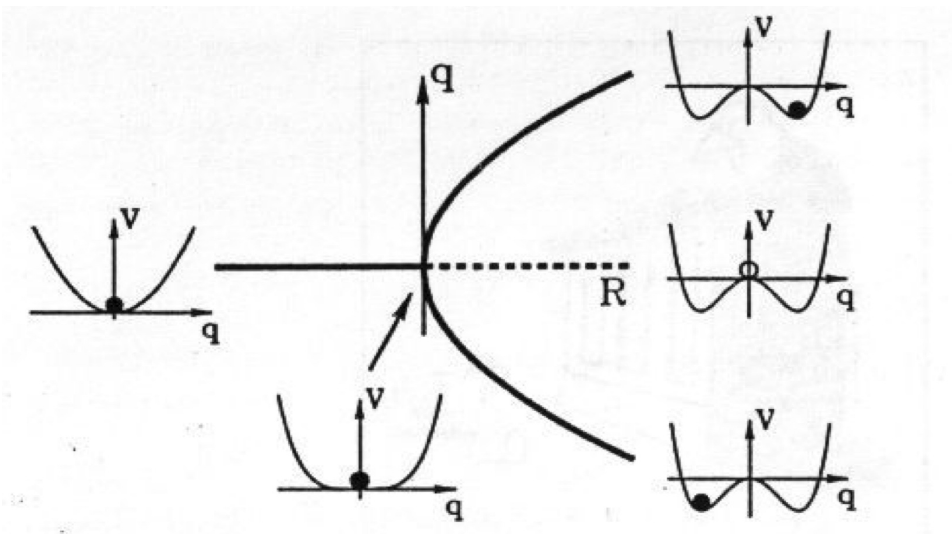

Fig. 14 A pitchfork bifurcation diagram and potential landscape for the emergence of convection rolls (adapted from Kelso 1995, p. 10)

Below the critical control parameter value, *R*, the fluid can be in only one macroscopic state, namely rest (Kelso 1995, p. 10). In figure 14, the diagram furthest to the left, the solid ball represents the state of the fluid as being in the minimum of its potential. *V* is the potential landscape. Therefore, the diagram furthest to the left represents *V*(q). The arrow represents an instability point (the open ball in the middle diagram to the right). At this stage a "stable solution of the dynamics becomes an unstable, and a transition to bistability occurs." (Kelso 1995, p. 11) After the bifurcation, "two rolling motions whose rotation speed is equal but opposite emerge spontaneously." (Kelso 1995, p. 10) Of course, only one particular rolling motion actually emerges (represented as either the bottom or the top diagram on the left side). If a set of generalized coordinates (to use Rueger's terms) together with *R* forms the subvenient base, and if the state of the fluid, *q*, expresses the supervenient properties of the system at a time prior to a bifurcation of the overall state, then it would seem that *V*(*q*) supervenes on a set of generalized coordinates together with *R* at a time prior to a bifurcation of the system.

However, as Rueger has set up his first argument, the *litmus test* for whether emergent properties supervene on a set of subvenient properties is whether or not *V* (the phase space portrait) remains qualitatively indifferent *after a bifurcation* occurs. Recall, according to Rueger, a system is structurally stable insofar as the following holds: if A*-properties supervene on A-properties at t_1, then a small change in A-properties – e.g., a small change in the value of the control parameter – will not result in a qualitatively different phase portrait at t_2

This condition, however, is violated in the example of convection rolls. First, at time t_1, say, the state of the fluid is in the macroscopic state of rest. Second, at time t_2, which in the diagram represents the onset of a dynamical instability point, the same physical properties at t_2 could be in two nomologically possible different states – either left or right rolling behavior. Finally, at time t_3, after the system has undergone a bifurcation, through a small change in *R*, temporal evolvement from t_1 to t_3 has led to the emergence of a qualitatively different phase space portrait. Prior to a relative small change in *R*, the state of the fluid was in the resting state, whereas after "wiggling" with *R*, the state of the fluid is in a continuously rising and falling state.

Before scrutinizing the second argument offered by Rueger, let us consider another problem with Rueger's first argument. Recall from chapter 3, that the standard way of distinguishing between synchronic and diachronic dependence relations is to conceptualize synchronic dependence as a *vertical dependence relation*, while conceiving of diachronic dependence as a *horizontal dependence relation*. That is precisely what Rueger does (see figure 13 above). That is, just as the usual account, the horizontal axis represents *time* or *change over time*, while the vertical axis represents *synchronic* or *atemporal* determination. The question I raised in chapter 3 was whether it is correct to insist so firmly on the distinction between horizontal and vertical, that is, between diachronic and synchronic modes of dependence.

Suppose that O'Connor (2000) is correct to claim that emergent properties in dynamical systems do not supervene on a set of physical properties, because what emergent properties a dynamical system has at t_1, e.g., is determined (in part, at least) by the immediately prior states of the system at t_0. If this account is correct, then it follows that the vertical synchronic relation does not exhaustively determine the emergent properties of the system at t_1. Rather, the relation of determination unfolds over time, from t_0 to t_1. As I argued in chapter 3, if this is the case, then the relevant relation of determination holds *diagonally* over time. But, if the relevant sense of determination holds diagonally over time, then this contravenes the requirement of the synchronic supervenience relation, since the latter can only hold – as Rueger himself states – *at a given synchronic time* (Rueger 2000, p. 472).

Turning now to look at Rueger's second argument for why diachronic emergence coincides with the supervenience relation, consider the following figure:

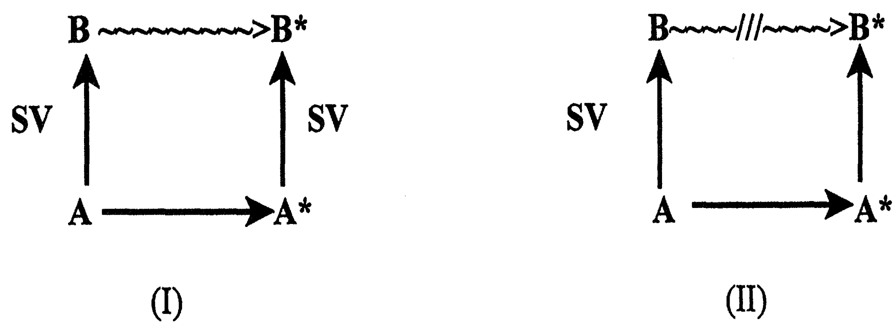

Fig. 15 Temporal evolution of a system {A, B} into {A*, B*}, where (I) the old and new systems are topologically equivalent (B* is a merely resultant property); in (II) old and new systems are inequivalent (B* is a diachronically emergent property); SV is the supervenience relation (adapted from Rueger 2000, p. 481)

The diagram to the left, (I), represents Rueger's first argument; here cashed out in terms of topological equivalence. Recall, if the phase space portrait stays qualitatively the same under perturbations, then the system is structurally stable, *viz*., that the trajectory of the old system {A, B} maps onto (that is, is equivalent to) the trajectories of the new system {A*, B*}[61].

The diagram to the right, (II), by contrast, is intended to show that when a perturbation of the system's dynamics occurs, the trajectories (behavior) of the old system {A, B} fail to map onto the trajectories of the new system {A*, B*}, in the sense that the new emergent behavior of the system {A*, B*} is qualitatively different and, as a result, topologically inequivalent to the old system {A, B}. Diagram (II) represents the *diachronic emergence* of novel behavior, e.g., the diachronic emergence of convection rolls. We know that if the trajectories of two physical systems are topologically inequivalent, then this violates the condition of *structural stability* required for emergent properties in *individual systems* to coincide with the supervenience relation. But, according to Rueger, instead of looking at topological inequivalence in individual systems, it is possible to define a conception of topological equivalence for *families of systems*, such that even though each individual system is structurally unstable, an individual system could nevertheless belong to a family of stable systems. As Rueger says: "If it does [belong to a family of stable systems], then the bifurcation or instability in the system occurs in a structurally stable way: all systems in the family exhibit qualitatively the same bifurcation behavior." (2000, p. 479) If so, then a supervenience relation exists between property classes in a system, if "the relation is unstable, the instability occurs in a stable way, i.e., the system belongs to a family of systems which, as a family, is structurally stable." (Rueger 2000, p. 479)

Suppose the supervenience relation holds between families of properties, and suppose further that dynamical systems such as those described in the convection roll example – which exhibit so-called Rayleigh-Bénard instability – belong to a family of systems in which qualitatively the same bifurcation behavior occurs,

---

[61] Rueger gives this the following formal definition: "A dynamical system Σ, considered as the transformations S_p: D → D on the system's phase space of initial conditions x at some times, is *topologically equivalent* to another system Σ* with S_{p*}: D* → D* if there exists a homeomorphism h (a one-to-one mapping continuous in both directions) of the phase space trajectories of the first system onto the trajectories of the second such that [(I)] commutes: h [S_{p*} (X)] = S_p [h (X)]." (2000, p. 473)

whenever these systems undergo a specific bifurcation, then it follows, on Rueger's account, that, as a family, Rayleigh-Bénard instability is structurally stable, and thus coincide with the supervenience relation. But, this result only follows under the specified condition that supervenience holds between classes or types of properties. The conclusion does not follow for *token systems*, where the trajectories of a system's phase space portrait at t_1 is topologically inequivalent with the trajectories of the new system's phase space portrait at t_2. On the one hand, then, we have a violation of the supervenience relation, whereas, on the other, the relation of supervenience remains intact. As a first approximation, this result is coherent, because when supervenience holds, it holds between types, and when it does not hold, it is because the relevant relation is between tokens.

This result implies that the relation of supervenience only holds in *generalized* cases across entire families of systems, while it does not hold in individual, non-generalized, systems. That is, claiming that diachronic ontological emergence takes the form of a supervenience relation, really boils down to saying that systems that exhibit diachronic ontological emergence do so because they belong to an *abstract* family of systems in which the relation of supervenience holds. It does not follow that any single *concrete* dynamical system that exhibits diachronic ontological emergent properties does so because such properties are based in a supervenience relation.

7.8. An implication for extended cognition

Given the arguments presented in sections 7.5.1, 7.6.1, and 7.7.1 – namely, that diachronic ontological emergence does not rely on the supervenience relation – I now wish to apply these arguments in order to consider their relevant implications for EC. Most defenders of EC conceive of extended cognitive systems as nonlinear dynamical systems. Arguments for such views are often grounded in the notion of *coupling* from dynamical systems theory. As Clark & Chalmers state, in cases of extended cognition, "the human organism is linked to an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right." (1998, p. 8) Such dynamical couplings may, in the right circumstances, result in the emergence of certain cognitive processes such as occurrent remembering. Consider, for example, how Clark defines his preferred view of emergence:

> "[The] idea is to depict emergence as the process by which complex, cyclic interactions give rise to stable and salient patterns of systemic behavior. […]. [P]henomena that depend on multiple, nonlinear, temporally asynchronous, positive feedback involving interactions will count as strongly emergent. […]. Emergent phenomena, thus defined, will […] reward understanding in terms of the changing values of a collective variable – a variable that tracks the pattern resulting from the interactions of multiple factors and forces. Such factors and forces may be wholly internal to the system or may include selected elements of the external environment." (2001, pp. 115-16)

Here the particular kind of dynamical coupling between agent and environment, say, involves complex, nonlinear, and self-organizing interactions from which certain emergent properties arise. That notion of emergence is, I submit, consistent with diachronic ontological emergence. In the literature, this specific kind of dynamics is often referred to as *continuous reciprocal causation*, which Clark defines as follows:

> "Continuous reciprocal causation (CRC) occurs when some system S is both continuously affecting and simultaneously being affected by, activity in some other system O. Internally, we may well confront such causal complexity in the brain since many neural areas are linked by both feedback and feedforward pathways […]. On a larger canvas, we often find processes of CRC that criss-cross brain, body and local environment. Think of a dancer, whose bodily orientation is continuously affecting and affected by her neural states, and whose movements are also influencing those of her partner, to whom she is continuously responding!" (1997, p. 163)

In addition to arguing for EC by way of dynamical notions such as coupling, through which one can identify instances of diachronic ontological emergence, one also finds appeals to supervenience as an attempt to ground the metaphysical claim of EC. Consider, again, a passage from Clark:

> "A recurrent theme in previous chapters has been the ability of body and world to act as what might now be dubbed "participant machinery" – that is, to form part of the very machinery by means of which mind and cognition are physically realized and hence to form part of the *local material supervenience base* for various mental states and processes." (2008, p. 207; italics added)

Given the arguments presented in this chapter, what should we make of this explicit appeal to supervenience? Insofar as Clark conceives of extended cognitive systems as dynamical or nonlinearly coupled systems, then it is not clear that it is justifiable to make use of both supervenience and emergence by which to ground EC. For purposes of explaining the dynamics of how complex and nonlinear dynamical systems give rise to higher-level emergent properties, the appeal to supervenience will do little work. Indeed, as Mitchell states:

> "If we take a snap-shot [synchronic] view of the higher and lower levels [in a dynamical system], then the dynamics of how the higher level is constituted and stabilized is lost. Contemporary sciences show us that there are processes, often involving negative and positive feedback or self-organization, that are responsible for generating higher-level stable properties, and these processes are not captured by a static mapping [such as supervenience]." (2012, p. 177)

That is wholly consistent with CRC and the notion of coupling from dynamical systems theory. The real culprit, it seems to me, is that when explaining emergent *processes*, and how these *arise* in *dynamical systems*, the supervenience relation fails to properly apply, because the supervenience relation is a synchronic relation, while, as we have seen, diachronically emergent properties are determined *diagonally over time*.

What I do not say is that there are no justified grounds for supervenience talk. For example, when Clark mentions that parts of the environment may, in the right circumstances, form part of the local material supervenience base of various mental states, perhaps he is correct to do so. I will not dispute this here. Nevertheless, I submit, the ramifications we *can* draw for EC given our discussion is the following: when addressing emergent cognitive processes, instantiated diachronically in nonlinear and self-organizing dynamical systems, the emergence of such processes refuse grounding in the supervenience relation.

7.9. Conclusion

In this chapter, I have explored the phenomenon of emergence. I have argued that insofar as we consider dynamical systems, and how the dynamics of such systems give rise to diachronic ontological emergent processes, there is no reason to assume that emergence in dynamical systems is coincident with supervenience. I have used my arguments for diachronic ontological emergence, together with my critical arguments of the supervenience relation, to address some implications for EC. That is, insofar as most defenders of EC understand extended cognitive systems as dynamical systems, and insofar as emergent processes in dynamical systems contravene the conditions under which the supervenience relation holds, friends of EC should avoid appealing to supervenience when explaining the emergence of cognitive processes.

8. Extended cognition & the causal-constitutive fallacy

In the story told so far I have criticized a number of metaphysical dependence relations for assuming ontological synchronicity. In contrast to synchronic views of metaphysical dependence relations, I have argued for the idea that metaphysical dependence relations need not be conceptualized synchronically but may, in the right circumstances, be understood to hold *diachronically*.

In this final chapter, the basic approach pursued in the thesis will remain the same, although here I examine what we might call a classic in the literature on EC, namely the *causal-constitutive fallacy* leveled against EC by some of its critics. Indeed, and for the first time in the thesis, it is now time to be entirely on the side of the EC paradigm, and defend EC against one of its central as well as recurrent metaphysical objections.

In the literature, the causal-constitutive fallacy has been widely discussed (see e.g., Adams & Aizawa 2001, 2008; Aizawa 2010; Clark 2008; Hurley 2010; Kirchhoff 2013a; 2013b; Menary 2006, 2007; Ross & Ladyman 2010; Wheeler 2010; to name but a few). However, it has typically been discussed in such a way to leave unquestioned the very concepts of constitution and causation (Hurley 2010; Kirchhoff 2013a; Ross & Ladyman 2010). In this chapter, I offer an analysis of the constitution relation, where I show that philosophical accounts of constitution have commonly been explicated in terms of synchronic relations between different entities. We are already familiar with this particular line of argument from chapter 3 and onwards, where we in chapter 3, for example, looked at how the relation of composition is standardly understood as a synchronic relation, albeit between several parts and a single whole. In line with the argumentational template I have used throughout the thesis, I then go on to show that Adams & Aizawa, the engineers behind the causal-constitutive fallacy in the EC literature, are using a synchronic notion of constitution to ground their universally quantified objection – the causal-constitutive fallacy – against EC. Such an application, I will argue, is problematic, because common examples of EC are implicitly consistent with a diachronic conception of constitution. I made this point back in chapter 1, where I mentioned that the diachronic view of dependence relations is implicit in a number of arguments provided for EC, especially in second-wave versions of EC. So, the causal-constitutive fallacy, I will argue, do not apply to common cases of EC. As a result, therefore, Adams & Aizawa are wrong to assume that their argument shows that EC *per se* is incoherent (as a matter of contingent fact).

8.1. Introduction

In metaphysics, the received view of constitution is known as "material constitution". Recall from chapter 2, where I provided a general exposition of constitution, that material constitution can be summarized as a synchronic one-one-relation between spatially and materially coincident objects of different kind, or as a many-one relation, where one object or entity is constituted by an aggregate of objects or entities (note: although the many-one expression is typically made in relation to composition, there are formulations of the constitution relation expressed in terms of many-one rather than one-one relations (see e.g., Wilson (2007)). In the context of material constitution, "synchronic" implies *atemporal* such that it is not part of the very essence of materially constituted entities that they are dependent for their existence on temporal unfolding. That is, material constitution involves a mode of constitution between entities in which time plays a role only insofar as it becomes possible to specify a constitutive relation at a time instant *t* – *viz*., moment by moment, snapshot by snapshot.

For example, as articulated by Gibbard (1975), in the classical case of *David* (a token statue) and *Piece* (a token piece of marble), both *David* and *Piece* are created at the exact same time and destroyed at the exact same time; hence, *David* did not evolve over time; rather, *David* is constituted at a time instant *t*. Here time is reduced to a set of punctuated specifications, with the fact that *David* is in time – just as *David* has a certain weight – not essential to the constitutive nature of *David*. Furthermore, together with the idea that constitution is atemporal, I emphasized, in chapter 2, that material constitution holds between two entities that spatially and materially coincide with one another. That is, if *Piece* constitutes *David*, both *Piece* and *David* exist at the same place at the same time and they share the same material parts.

This chapter offers an alternative to this view of constitution by focusing on the dispute between defenders of EC and their critics. As we know, advocates of EC state that in orchestration with neural elements, extra-neural bodily, and worldly elements partially *constitute* cases of distributed cognitive processes or modes of cognitive processing (Clark & Chalmers 1998; Clark 2008; Menary 2007; Sutton 2010; Wheeler 2010; Wilson & Clark 2009; Wilson 2010).

But, there are philosophers who claim that the defenders of EC commit the so-called "causal-constitutive fallacy" (C-C fallacy) (Adams & Aizawa 2001, 2008; Aizawa 2010). Because issues such as causal coupling are part of the argument for EC, those who think that EC commits the C-C fallacy are arguing that defenders of EC make an unjustifiable inference from causal dependence to constitutive dependence. The reason for this is that it is commonly thought that causation and constitution are independent relations such that facts about causal relations do not tell us anything about facts of constitutive relations. Hence, so the critics argue, nothing follows about constitution from facts about causation on its own.

As a first approximation, and if material constitution is the received view of constitution, it may seem that the critics of EC are correct in charging the defenders of EC with the C-C fallacy. If we consult the case of *David* and *Piece*, it makes little sense to start with claims about causation and then infer to claims about constitution – *Piece* constitutes *David*, *Piece* does not cause *David* to exist. But first approximations are not always accurate. Indeed, it only seems that those who argue that the fans of EC are guilty of committing the C-C fallacy are correct, because those critics have misunderstood the nature of the constitution relation involved in common examples of extended cognitive processes or modes of cognitive processing[62].

[62] Another critic of the extended cognition framework is Rupert (2009). I omit discussion of Rupert in this chapter, because Rupert's work, even though it takes the metaphysics of mind seriously, is focused on integration in terms of mechanisms

If I am correct to insist on the need for an alternative conception of constitution, this would have ramifications for the metaphysics of constitution, in that, it demonstrates the need to broaden how we conceive of the constitution relation itself, on the one hand, and what kinds of relata the constitution relation may hold between, on the other. This should not come as a surprise to us after the discussion of how *processes* depend for their persistence on spatiotemporal continuity in chapters 3, 4, and 5. Also, the need for an alternative conception of constitution points to the failure of some philosophers to pay attention to the metaphysical baggage their statements carry with them, and, consequently, involve them in. Although the issue that I shall argue in this chapter starts within naturalistic philosophy of mind (the extended cognition thesis), it ultimately speaks to wider issues about constitution in analytical metaphysics. Consequently, and in line with one of my primary aims in the thesis, by joining up distinct-ish literatures in analytical metaphysics (on constitution) and in naturalistic philosophy of mind (on the extended cognition thesis), it is my hope to get audiences from both fields of research to contribute to one another's projects (as shown in figure 1).

8.2. Arguments

The issue that I shall discuss in this chapter is that there is an assumption shared on both sides of the EC debate that is misleading and that requires the development of an alternative, nonstandard account of the nature of constitution for it to be resolved. I identify, first, what the misleading assumption is in the EC literature. What is misleading is the assumption that the constitution relation, in distributed cognitive processes or modes of cognitive processing, is (a) *synchronic* and (b) *fundamentally distinct* from causation. This assumption generates two interesting implications. First, as a universally quantified argument against all cases of EC, I will argue that the C-C fallacy turns on an argument that is wrong, namely that the conception of constitution used by the defenders of EC must be compatible with how it is characterized in analytical metaphysics. However, common examples of EC do not dovetail with the notion of constitution developed in metaphysics, and

---

rather than on the constitution relation, the latter being the topic of discussion in this chapter.

vice versa. So, critics of EC such as Adams & Aizawa are wrong to insist that the C-C fallacy points to something flawed in all cases of EC. Second, even if common cases of EC are incompatible with the standard account of constitution in metaphysics, some defenders of EC (e.g., Clark 2008; Rowlands 2010), I will argue, misleadingly adopt metaphysical concepts such as constitution without additional explanation and scrutiny, thus (potentially) misconstruing the relation of constitution in cases of distributed species of cognitive processes.

As I mentioned above, the diachronic conception of constitution to be developed here is (in fact) implicit in some articulations of EC – especially in second-wave versions of EC (see e.g. Menary 2007) – even if this diachronic notion of constitution has not been explicitly articulated. Thus, part of my project is to make explicit what is already in the existing literature on EC. However, the diachronic account of constitution on offer here undermines other views of constitution that are equally held in the debate over EC – that is why I state that Adams & Aizawa are wrong in charging all cases of EC with the C-C fallacy, since the claim they advance presupposes the applicability of synchronic material constitution to EC. Furthermore, that is why I say that certain defenders of EC (e.g., Clark 2008; Rowlands 2010) are wrong to adopt metaphysical relations from metaphysics without additional scrutiny.

To extend on these remarks, this chapter offers a challenge to those who have either argued from material constitution to the claim that EC commits the C-C fallacy (which I will show that Adams & Aizawa have) or thought it plausible to base an argument for EC by appeal to constitution as a synchronic relation of dependence between different entities (which I will show that Clark (2008) and Rowlands (2010) have), while attempting to articulate a diachronic conception of constitution that I think is present in certain formulations of EC.

For instance, in section 8.5, I refer to Wilson's recent review of the metaphysical literature on constitution (2007, 2009) in order to make the claim that the notion of constitution that Adams & Aizawa (2001, 2008) are working with is what Wilson identifies as *compositional* constitution – a species of material constitution used in metaphysics. This specification is important, since (as I shall argue) this notion of material constitution is incompatible with the constitutive nature of distributed cognitive processes. As a result, I argue that Adams & Aizawa are wrong to charge all advocates of EC with the C-C fallacy. Wilson also identifies a second species of material constitution, which he refers to as *ampliative* constitution. As with compositional constitution, I shall spend some time substantiating the claim that this second form of material constitution is equally inconsistent with common cases of EC.

The alternative to what I shall refer to as *synchronic material constitution*, which covers both compositional constitution and ampliative constitution, I call *diachronic constitution*. To get an initial fix on what this nonstandard mode of constitution is, and how it is compatible with most of the constitutive cases of EC, I shall start by identifying that extended cognitive phenomena are hybrid, that is, they exhibit both causal and constitutive relations. However, even though distributed processes are hybrid in this sense, I shall question the assumption inherent in the standard notion of constitution, that is, whether it is justified to understand constitution as a strictly vertical relation, while understanding causation as an exclusively horizontal relation. I introduced a similar worry in chapter 3 and 5, where I made the argument that insofar as the composition relation may hold diachronically, it follows that this kind of dependence relation holds *diagonally* across time rather than vertically at a durationless point in time.

Here I argue that one is justified in claiming that the constitution relation that holds between *Piece* and *David* at t_1 does so in virtue of the fact that both *Piece* and *David* are wholly present at t_1, whereas it is due to certain diachronic or causal relations that *Piece* and *David* persist from t_1 to t_2. However, if the relata in question are *processes*, and because processes are temporally extended in time, the constitution relation cannot hold exhaustively between its relata at any single moment in time but must do so *across time*.

Furthermore, and as I argued in chapters 3 and 5, processes – whether cognitive or non-cognitive – persist by perduring rather than by enduring. This point is closely related to why constitution cannot hold in a vertical fashion but must hold diagonally in circumstances when the relata in question are processes. Consider, for example, the following case by Menary: "*X* is the manipulation of the notebook *reciprocally* coupled to *Y* – the brain processes – which together constitute *Z*, the process of remembering." (2006, p. 334; italics in original) In this example, *Z* is a temporally extended process – just as *X* and *Y* are. That is, *Z* persists through time in virtue of its spatiotemporal continuity (Hofweber & Velleman 2011), and at no single (snapshot) instant in time is *Z* exhaustively present (see chapter 5 for a similar argument in the case of transactive remembering). In other words, *Z* does not persist by being wholly present at each time instant *t* at which *Z* exists. Indeed, Menary provides us with a temporally multifaceted example of EC. That is, the timescales that are involved include the time-courses of neural processes, temporal dynamics of bodily manipulation, and the temporal dynamics of cultural practices within which the overall distributed process unfolds. Also, the time-course of *Z* is longer than some of its components, e.g., *Z* has a longer time-course than certain top-down neural modulations involved in orchestrating *Z* (see chapter 6). Thus, *Z* is constituted by processes unfolding over different temporal frequencies (milliseconds, seconds, and minutes), while *Z* itself is dynamically unfolding over a time interval. That is why neither the constitution relation that holds between *X* & *Y* and *Z* nor any of the relata can be wholly present at a particular instant *t* or at each point over an interval $t_1, \ldots, t_n$.

What this tells us about the relation of constitution in EC is that relations of constitution do not sit synchronically or atemporally wedged in between higher and lower-level entities as in cases of material constitution. Unlike the notion of time expressed in synchronic material constitution, where "time" is reduced to a series of snapshot instances, and where "time" is not essential to the nature of the constitu*ted*, distributed cognitive processes are constituted in *time continuous* dynamical systems, and dynamical systems are *quantitative in time* (van Gelder 1998, p. 618; see also Chemero 2009). By "quantitative in time," I mean that both the constituents and the constituted are richly embedded in time such that if we change this embedding – e.g., by slowing down some of the processes – we change either the behavior of the lower level processes and their components or the higher level phenomenon, or both, depending on which of the processes we change (Smithers 1998, p. 652). As Wheeler points out, in the psychological arena, richly temporal phenomena include:

> "(i) [The] rates of change within, the actual temporal duration of, and any rhythmic properties exhibited by, individual cognitive processes, and (ii) the ways in which those rates of change, temporal durations, and rhythms are synchronized both with the corresponding temporal phenomena exhibited by other cognitive processes, and the temporal processes taking place in the cognizer's body and environment." (2005, p. 106)

If the notion of "time" in synchronic accounts of material constitution implies that temporality itself is not essential to the constitutive nature of constituted entities, then the metaphysical tool-kit of material constitution will be inappropriate for describing and explaining dynamical systems, and the way in which such systems give rise to distributed cognitive processes.

8.3. Overview

In section 8.4, the task will be to consider a few examples to serve as a backdrop for the discussion that follows in later sections. In section 8.5, I review Wilson's two senses of material constitution, while (a) showing that Adams & Aizawa are working with the notion of material constitution known as compositional constitution, and (b) develop the argument that whether adopting either compositional or ampliative constitution is irrelevant, since neither is compatible with cases of distributed cognitive processes. I also develop my second argument, namely that Clark (2008) and Rowlands (2010) seem to misleadingly adopt synchronic notions of constitution when they should, in fact, be appealing to diachronic forms of constitution (or, of course, simply not use such metaphysical concepts at all). Now, having done this, I begin to contrast the concept of synchronic material constitution with diachronic constitution in relation to the C-C fallacy in section 8.6. Finally, in section 8.7, I focus on developing further the idea of diachronic constitution.

8.4. Examples

A first task is to sketch and discuss a few examples to serve as a backdrop for the discussion that will follow in later sections. I begin with the critics of EC. In his "The coupling-constitution fallacy revisited" (2010), Aizawa says: "Once one sees that a causal connection between a process of type X and a process of type Y is not enough to convert the Y process, or even the conjoined X-Y process, into a process of just type X, then one can also see that essentially the same point applies even when there is a reliable causal connection between X and Y." (2010, p. 333) According to Aizawa, there are persistent intuitions to the effect that those who infer constitution from causation are committing an instance of the C-C fallacy. As one of Adams & Aizawa's leading cases to illustrate this argument, they provide this non-cognitive example:

> "The liquid Freon™ in an older model air conditioning system evaporates in the system's evaporator coil. The evaporator coil, however, is causally linked to such things as a compressor, expansion valve, and air conditioning ductwork. Yet, the evaporation does not extend beyond the bounds of the Freon™. So a process may actively interact with its environment, but this does not mean that it extends into its environment." (2008, p. 91)

Regardless of whether one finds the description "evaporation does not extend beyond the bounds of the Freon™" a bit pseudo-scientific (Ross & Ladyman 2010, p. 161), Adams & Aizawa are, of course, correct to claim the following: "a compressor and an expansion coil have complementary roles to play in air conditioning, although this provides no reason to think that a compressor is an expansion coil or vice versa." (2008, p. x) Indeed, most central air conditioners have two separate components: the first is a condenser coil; the second is an evaporator coil. The evaporator is typically mounted inside the house. It is an inner coil in a heat pump that, during cooling mode, absorbs heat from inside the house and boils the liquid refrigerant (e.g., Freon™) to a vapor, which then cools the house. In contrast, the condenser is typically placed outside the house. It is a network of tubes filled with refrigerant that removes heat from the hot evaporated liquid so that the refrigerant becomes a liquid again.

There is a serious problem with employing such an example as evidence for the unjustifiability of putative cases of EC. As I have mentioned, extended cognitive phenomena are hybrid, consisting of both causal relations and relations of constitution. Constitution is usually understood as a relation between levels, and is therefore typically conceived as an interlevel relation of dependence. In contrast, the relation of causation is commonly taken to be a relation that holds at the same level, and is therefore usually viewed as an intralevel relation[63]. In this example, Adams &

[63] I should add here that I do not restrict causation to an intra-level phenomenon. In chapter 7, for example, I explicit talk of ontological emergence that involves downward causation.

Aizawa conflate intralevel causal relations with interlevel constitutive relations, in the sense that they start by assuming that the evaporator coil and the condenser coil are in constant causal interaction and, then, conclude from this that neither the evaporator coil nor the condenser coil is constitutive of one another. If Adams & Aizawa insist that their example works against EC, they would be deliberately twisting the interpretation of cases of EC.

For simplicity only, consider again the relationship between *Z* (the distributed process of remembering) and its constituents *X* and *Y* (the process of manipulating a notebook causally coupled to brain processes). By analogy, if the example of evaporation in an air conditioning system were to map on to the example of distributed remembering, it would follow that defenders of EC were arguing such implausible things as: that the manipulation of a notebook extends into the brain processes, and, therefore, that *X* (the manipulation of the notebook) is *Y* (various brain processes). Unsurprisingly, no one on the side of EC has ever made such a remarkably strange claim (at least not anybody that I know of). It would be consistent with most cases of distributed cognitive processes, if Adams & Aizawa were to reformulate their example accordingly: *X* is the process of absorbing heat and thus transforming a liquid refrigerant into vapor (the function of the evaporation coil) reciprocally coupled to *Y* – the process of transforming the vapor into liquid refrigerant (the function of the condenser coil) – which together constitute *Z*, the process of maintaining constant room temperature. This reformulation would respect the distinction between causation and constitution, with causation occurring at the constituent level, and constitution holding between the constituted higher level and the constituents at lower levels.

For an example of a different kind, albeit still non-cognitive, consider these points concerning the familiar case from dynamical systems theoretical approaches to cognition: the Watt (centrifugal) governor[64]. First, the Watt governor is a mechanism for controlling the speed of a steam engine. Second, Watt solved the problem of maintaining constant speed for the flywheel by attaching a vertical

[64] In the article "What might cognition be, if not computational?" (1995), van Gelder recommends that the operations of the Watt Governor, along with a dynamical mathematical description of its operations, be understood as a prototypical model for cognitive science and for the ontology of cognition.

spindle to the flywheel, which would rotate at a speed proportionate to that of the flywheel itself. Watt then attached two arms with metal balls on their ends to the spindle; both were free to rise and fall and, as a consequence of centrifugal force, would do so in accordance with the speed of the governor. Due to a mechanical arrangement, the angle of the arms would change the opening of the valve, thus having an effect on the amount of steam driving the flywheel. This provided a system, the result of which was "that as the speed of the main wheel increased, the arms raised, closing the valve and restricting the flow of steam: as the speed decreased, the arms fell, opening the valve and allowing more steam to flow. The engine adopted a constant speed, maintained with extraordinary swiftness and smoothness in the presence of large fluctuations in pressure and load." (van Gelder 1995, p. 349)

Standardly the example of the Watt governor (WG) has been used as a prototype of a dynamical system with the grand ambition of establishing a non-computational and non-representational approach to cognition and cognitive science. For the present purpose, however, I bracket these agendas. Instead I use this case to further motivate the idea of diachronic constitution between dynamically evolving processes. Similarly to cases of EC, the centrifugal governor is hybrid, that is, it exemplifies both relations of constitution and causation. Even if talk of causation might seem problematic in cases where there is a constitutive relation between higher- and lower-level relata (Craver & Bechtel 2007), this is unproblematic here. That is, by attending to the constitutive relationship between *Y* (the process of maintaining a constant speed level) and the sub-processes such as closing of the valve, the rotation as well as height of the spindle arms, etc., we can trace both intralevel processes of reciprocal, mutually influencing causation between lower-level processes and their components, while explicating what is the constitutive relation between *Y* and its constituents. Because of this, the correct way to *causally* explain how the centrifugal governor works is to explain the mutually modulatory and interconnected character of the components, whereas the engine speed is *constituted* by the overall dynamics of this lower-level activity.

Before furthering my analysis of how to understand the distinction between synchronic material constitution and diachronic constitution, there is an issue that needs attention. To begin, then, note that in discussions of constitution, what we are presupposing is that we are discussing the *nature* of some phenomenon, *viz*., that phenomenon's ontology. This, however, poses an important question, namely whether insights from nonlinear dynamics provide us with an account of *epistemological* constitution, where the constituted phenomenon is merely some artifact of a particular model or formalism generated through macroscopic analysis, whether we get an account of *ontological* constitution, which informs us about the *nature* of the system in question, or if we get an account of constitution that straddles both categories? In section 8.7.5, I show that dynamical systems are ontologically constituted, in that, the relation between higher- and lower-level processes is such that it exhibits both bottom-up and top-down *constitutively mediated effects* (Craver & Bechtel 2007; or, "downward causation" (see chapter 7)). Even if all dynamical systems are instances of what I have termed "epistemological constitution," because modeling is necessary in order to understand the behavior of the system – as its interdependent variables unfold over time – dynamical systems are also ontologically constituted. This justifies my use of systems like the WG in the debate about constitution in EC.

To continue distinguishing between cases of synchronic material constitution and the test case of diachronic constitution (Watt's governor), unlike the example of *David* and *Piece*, the relation of constitution in the centrifugal governor does not consist in a relation between enduring entities that are wholly present whenever they exist. Rather, in the centrifugal governor, the relation of constitution holds between processes at both higher- and lower-levels. To persist, in the sense of processes, is to persist by unfolding over time, while never being wholly present at any single instant *t*. I have made much of this point already in the preceding chapters, especially in chapters 3, 4, 5, and 6, where we looked at the nature of the relata involved in a Mexican wave and writing a cheque (chapter 3), the process of CA (chapter 4), transactive remembering (chapter 5), and free energy minimization (chapter 6). Even though it may be a conceptual truth (Hofweber & Velleman 2011) that processes are temporally extended such that they require spatiotemporal continuity for their persistence, it gives rise to a substantial question: how can the relation of constitution that holds between processes be atemporal (i.e., synchronic) if the relata themselves (at both higher- and lower-levels) are *essentially* temporal (i.e., diachronic)? In other words, if the very nature of a process is to unfold over time such that it can never be wholly present at any single instant *t*, how, then, can its existence be exhaustively determined at a synchronic instant?

In the case of *David* and *Piece*, where the constitution relation is understood to hold entirely between enduring entities, which are wholly present whenever they exist, there seems to be no problem with claiming that the constitution relation that holds between *David* and *Piece* is synchronic. However, insofar as the relata do not persist by enduring, it is problematic to insist that the higher-level relatum is constituted synchronically – both in the case of the centrifugal governor and Menary's case of distributed remembering.

Of course, even in dynamical cases like the Watt governor, the constitution relation between the higher-level process of maintaining a constant speed and the relevant sub-processes and their components at a lower level may appear to hold synchronically. But, I submit, this is only because of our need to represent it spatially by drawing, for example, a representation of the dynamical system on a blackboard or depicting the operations of the Watt governor in a static representation. However, this spatial representation is misleading; the spatial representation does not map onto the dynamical characteristics of the constitution relation when we consider in detail the dynamical evolution inherent in the centrifugal governor.

Thus, specifying the temporal dynamics of higher-level processes, along with the temporal nature of lower-level processes, and explaining the constitution relation between these, is what I aim to do by pushing for the concept of diachronic constitution. This proposal obviously conflicts with metaphysical intuitions about constitution. But this conflict is both unavoidable and necessary if we are to explain the constitutive nature of distributed cognitive processes.

8.5. Two concepts of synchronic material constitution

A critical step in articulating my first argument – that the C-C fallacy is wrong and thus misleading – is to establish what notion of constitution Adams & Aizawa are working with. As I have already claimed, Adams & Aizawa are working with the specific notion of compositional constitution, a species of what I call synchronic material constitution. This specification I now need to spell out in more detail. To do so, I piggyback on Wilson's recent review of the metaphysical literature on material constitution (2007, 2009), where Wilson distinguishes between compositional and ampliative constitution, both of which are modes of synchronic material constitution. The benefit of surveying this review of material constitution is that it allows me to argue that regardless of Adams & Aizawa taking either compositional constitution or ampliative constitution onboard – from which to justify their claim that EC commits the C-C fallacy – neither of these can justify such a critical argument.

Both compositional and ampliative constitution share two necessary conditions that any adequate analysis of synchronic material constitution must satisfy: *y* is materially constituted by *x* (or the *xs*) during p only if:

> "**Coincidence**: *x* is completely material in itself, or the *xs* are completely material in themselves, and *y* is spatially and materially coincident with *x* (or the *xs*) during p." (Wilson 2009, p. 370)

> "**Distinctness**: it is possible for *x* (or the *xs*) to exist without there being anything of *y's* type that is (even partially) spatially and materially coincident with *x* (the *xs*). (Wilson 2009, p. 370)[65]

These two standard conditions of material constitution motivate most of Adams & Aizawa's intuitions about the C-C fallacy, I submit. With the case of the air conditioning system as our backdrop, consider that the coincidence condition requires a specific form of *overlap*, namely the *complete* overlap of space and material between *y* and *x* (or the *xs*) for the duration of the constitution relation (Wilson 2007, p. 5). For instance, it is coherent to say that two or more roads *partially* overlap, since at their spatial intersection, they share the same material. Similarly with the case of the air conditioning system, where it seems equally coherent to say that the condenser coil and the evaporator coil share parts of the same material. However, just as two or more overlapping roads do not share the same spatial and material parts completely, the evaporator coil and the condenser coil do not share the same spatial and material parts completely. Indeed, if this is a necessary condition that all adequate analyses of material constitution must fulfill, it

[65] I add the distinctness condition here for completeness. While I mention it one more time in the discussion to come, my real interest is with the coincidence condition.

is easy to see why Adams & Aizawa make the claim that the evaporator coil does not constitute the condenser coil, and *vice versa*.

With this condition explained (although it can be done so in more detail), we can already see why Adams & Aizawa are mistaken in arguing that all cases of EC commit the C-C fallacy. Recall the schematically defined case of EC by Menary: "*X* is the manipulation of the notebook *reciprocally* coupled to *Y* – the brain processes – which together constitute *Z*, the process of remembering." (2006, p. 334; italics in original) In this example, *X* and *Y* refer to lower-level processes, while *Z* refers to a higher-level process. Keep this in mind so as not to be confused by my use of X and Y above in the example given by Wilson, and my use of X and Y here in the example by Menary.

Once we have this picture – that *X* and *Y* combine to jointly constitute *Z* – it is clear why the coincidence condition of synchronic material constitution cannot be employed to justify the argument that all defenders of EC commit the fallacy of conflating causation with constitution. On the one hand, nowhere does Menary claim that *X* and *Y* must spatially and materially coincide completely. Indeed, this would be a rather bizarre thing to claim. On the other hand, even if it is possible for *X* and *Y* to exist without there being anything of *Z's* type (this is the requirement of non-identity stated in the distinctness condition), *Z* does not spatially and materially overlap completely with *X* and *Y*, and *vice versa*. I find it fully coherent to claim that the relationship between *X* & *Y* and *Z* is such that the space-time path of *Z* includes the space-time path of *X* & *Y* (this corresponds to the definition of composition provided in chapters 2 and 5). This is because, if *X* & *Y* constitute *Z*, higher-level processes such as *Z* is built up from *X* & *Y*. However, it does not follow from this that a higher-level process (like *Z*) and its lower-level constituent sub-processes (like *X* & *Y*) completely overlap materially with one another. That *X & Y*, on the one hand, and *Z*, on the other, do not completely overlap materially with one another can be illustrated by highlighting that *Z* and *X & Y* are embedded dynamically at multiple different timescales. That is, neural assemblies run over timescales of milliseconds, whereas the practice of manipulating a notebook runs over longer timescales. Thus, it makes little sense to insist on material coincidence given the dynamical nature of distributed processes, like *Z*, and their sub-processes and components, like *X* and *Y*.

8.5.1. Compositional constitution

To continue the analysis of material constitution, what Wilson dubs *compositional* constitution, has, in addition to the conditions of coincidence and distinctness, two further necessary conditions that it is commonly expressed to imply:

> "**Intrinsic Necessitation**: *x* is in some intrinsic state(s), or the *xs* that compose *y* are arranged, during p such that *x* itself, or the *xs* themselves, necessitate the existence of *y*." (2009, p. 371)

> "**Constituent Necessitation**: whenever *y* exists, there must be something of *x's* type that is […] spatially and materially coincident with *y*." (2009, p. 371)

In the intrinsic necessitation condition the idea that constitution is a compositional, part-whole, relation finds its most clear expression. It is now possible to address my second argument, namely that certain defenders of EC take for granted the applicability of metaphysical notions such as constitution without taking the additional step of providing a proper analysis of such terms. Consider, for example, these two quotes by Clark (2008) and Rowlands (2009) respectively:

> "We thus come to what is arguably the most radical contemporary take on the potential cognitive role of nonbiological props, aids, and structures: the idea that, under certain conditions, such props and structures might count as *proper parts of extended cognitive processes*." (Clark 2008, p. 68; italics in original);

> "*EM* is a claim about the *composition* or *constitution* of (some) mental processes." (Rowlands 2009, p. 54; italics added; see also Rowlands 2010)

Both authors express a commitment to constitution as compositional – Rowlands does so explicitly, while Clark does so by using the notion of "proper part," which refers to the formal theory of extensional mereology.

It is misleading for defenders of EC to make use of compositional constitution for several reasons. First, compositional constitution is a strict partial ordering, like the notion of a "proper part" in the formal theory of extensional mereology. A partial ordering is a reflexive, transitive, and antisymmetric relation (Varzi 2009, p. 4). In the context of extensional mereology, we can understand transitivity as stating that "any part of any part of a thing is itself part of that thing." (Varzi 2009, p. 4) Formally expressed: $((Pxy \wedge Pyz) \rightarrow Pxz)$. That a partial ordering is antisymmetric can be expressed as follows: "Two distinct things cannot be part of each other." (Varzi 2009, p. 4) Formally expressed: $((Pxy \wedge Pyx) \rightarrow x = y)$. Finally, reflexivity implies that "[e]verything is part of itself." (Varzi 2009, p. 4) This is formally expressed as follows: Pxx. There are individual problems with each of these conditions. For example, as I signposted in chapter 2, the view that constitution is transitive is controversial, because one can observe legitimate senses of "part" that are not transitive. Consider, e.g., these two arguments:

1.A. This chain is constituted by metal links.
1.B. Those metal links are constituted by physical particles.
1.C. This chain is constituted by physical particles.

In this case, the premises (1.A) and (1.B), together with the criterion of transitivity, entail (1.C). But what about the following argument:

2.A. This queue is constituted by a sequential order of people.
2.B. That sequential order of people is constituted by physical particles.
2.C. This queue is constituted by physical particles.

This argument appears to have the same form as (1A-1C). However, even if both arguments rely on transitivity, unlike (1A-1C), (2A-2C) is controversial, in that, it is not clear that (2A-2C) can accommodate transitivity. Specifically, unlike a metal chain, which one might think of as nothing more than various entities appropriately organized, queues are more than simply their physical parts – regardless of how these might be arranged. According to Wilson, if this is correct, then there is a non-trivial metaphysical difference between entities such as a chain and a queue and their constituents. As Wilson puts it: "Consider any chunk of physical matter. If you merely add physical matter to this chunk, there will be a way to do so that itself creates a chain. But there is *no* such way of proceeding here that creates a statue."

(2009, p. 369; italics in original) The same is true of a queue. Mere addition, or, for that matter, arrangement, will not suffice to constitute a statue or a queue, since they are partly individuated by physical, intentional, and socio-cultural relations "that pertain in the broader locale of that constituent physical matter." (Wilson 2009, p. 369) We can simply note that not all sequential line configurations of people constitute a queue (Hutchins 2005, p. 1559). Soldiers standing at attention are in a sequential line configuration, yet they do not constitute a queue. Part of what makes a sequential line configuration a queue is that that sequential line configuration is embedded within cultural practices with appropriate norms (Hutchins 2011a).

The second problematic element concerning compositional constitution being a strict partial ordering – insofar as some defenders of EC use this notion of constitution – is that the principle of reflexivity states that everything is a part of itself. But it is (*prima facie*, at least) counter-intuitive to view entities like people, in our argument (2A-2C), as parts of themselves (Wilson 2007, p. 7). If it is counter-intuitive to view people – or objects such as *David* and *Piece* – as parts of themselves, then it is also (*prima facie*, at least) counter-intuitive to count a process (distributed remembering, say) as a part of itself. Finally, both Clark and Rowlands emphasize that they focus on *processes* in the above quotes. This, however, since compositional constitution is a synchronic relation of dependence, and processes are diachronic in their very essence, provides Clark and Rowlands with the dilemma I outlined earlier: how can the relation of constitution that holds between processes be atemporal (i.e., synchronic) if the relata themselves (at both higher- and lower-levels) are essentially temporal (i.e., diachronic)? In other words, if the very nature of a process is to unfold over time for a process to *be* what it is, then how can its existence be determined at an atemporal instant? The answer, we have seen throughout the thesis, is that it cannot. Therefore, Clark and Rowlands ought to avoid the employment of compositional constitution to argue for EC[66].

[66] In the event that either Clark or Rowlands, or both, claim that they do not employ constitution in the form that I have just argued that they do, the ramifications of my argument shift from pointing out that they are mistaken in such an employment to providing an articulation of what they implicitly endorse. Either way, there are gains to be made from this discussion.

8.5.2. Ampliative constitution

The second concept of synchronic material constitution identified by Wilson (2007, 2009) is *ampliative* constitution. This notion is interesting to consider simply because it is not characterized by intrinsic necessitation conditions but rather by two conditions that go beyond relations of extensional, part-whole, mereology to consider *contextual* and *relational* aspects of both the constituents as well as the constituted phenomenon. One might think that because of its relational and contextual aspects that this relation of constitution is precisely the kind of constitution relation defenders of EC should be working with. This can be motivated further, since the mode of ampliative constitution explains the underlying intuition that the constituted phenomenon is *more than simply* its internal physical constituents – that is, there is more to the constitution of a phenomenon than is physically "within" that phenomenon.

Unfortunately we have to do much better than appealing to ampliative constitution in order to explain the mode of constitution in distributed cognitive processes. To get a fix on this, I first need to highlight and explain what ampliative constitution is.

Let us begin by noticing, as Wilson does, that the concept of ampliative constitution has been particularly important in the work of Baker (1999, 2000) – with the aim of establishing a constitutive view of persons. For my purposes here, the discussion of persons is irrelevant. Instead I shall keep my focus on the example introduced at the beginning of this chapter, the relation between *David* and *Piece*. For Baker, "x constitutes y at $t =_{df}$

a) x and y are spatially coincident at t and share all the same material parts at t; and
b) x is in D at t; and
c) It is necessary that $\forall z$ [F* zt and z is in D at t) $\rightarrow \exists u$(G^* ut and u is spatially coincident with z at t)]; and
d) It is possible that (x exists at t and $\sim\exists w$[G^* wt and w is spatially coincident with x at t]); and

e) If *y* has any nonspatial parts at *t*, then *x* has the same nonspatial parts at *t*." (1999, p. 149; see also Baker 2000)

As Wilson (2009) has shown, the case of *David* and *Piece* satisfies conditions (a)-(e). *David* and *Piece* are both materially as well as spatially coincident during *t*. *Piece* is in art-rich cultural circumstances, given a title, and put on display at *t*. Necessarily for anything that has "being a piece of marble" as its modal property (i.e., the mode in which a property is had necessarily or possibly by an object) and is presented as a figure in art-rich cultural circumstances, given a title, and put on display at *t*, then something exists that has "being a statue" as its modal property that materially and spatially coincides with "being a piece of marble" at *t*. In (d), the modal claim made is that it is possible that *Piece* exists at *t* and that *Piece* does not spatially and materially coincide with anything that has "being a statue" as its modal property at *t* (*Piece* and *David* differ with respect to their persistence conditions).

According to Wilson, the view of ampliative constitution is based on the two following conditions:

> "**Extrinsic necessitation**: *x* (the *xs*) is (are) in extrinsic conditions during p that themselves necessitate the existence of *y*." (Wilson 2009, p. 371)
>
> "**Relational/Intrinsic Constraint**: *y* is relationally individuated and *x* (the *xs*) intrinsically individuated." (Wilson 2009, p. 371)

With these two conditions defined, we can now assess which notion of constitution Adams & Aizawa employ: whether it is compositional constitution; or, whether it is ampliative constitution. Since I have already argued that Adams & Aizawa assume the soundness of compositional constitution, and that this notion of constitution is inconsistent with distributed cognitive processes, I focus on the second notion here. I will show that ampliative constitution is equally problematic.

First, if we attend only to the relational/intrinsic constraint of ampliative constitution, this specific constraint is insufficient twofold. It is insufficient for the defender of EC to attempt to ground the constitution claim of EC. The fact that a process *Z* (e.g., the distributed process of remembering) is relationally individuated will not suffice to establish the claim that environmental elements play a constitutive role in certain tokens of cognitive processes. Put differently, it is not enough to show that certain socio-cultural circumstances are causally necessary for a process to be extended or not. In fact, in contrast to EC, where some of the physical constituents of a cognitive process may be located "outside" the brain and body of an individual, the relational/intrinsic constraint specifies that all the physical constituents are located *internally* to the individual in question. Moreover, if Adams & Aizawa take this particular form of synchronic material constitution onboard, it will not be possible for them to justifiably underpin the charge that defenders of EC commit the C-C fallacy. Instead, such an argument would beg the question against the relation of constitution in EC. If the constitutive nature of distributed cognitive processes does not reflect the kind of constitutive character implied by relations of ampliative constitution, then how could critics of EC, like Adams & Aizawa, base the argument for the C-C fallacy on ampliative constitution? Indeed, they could not.

Second, in his recent review of material constitution, Wasserman (2009) discusses four different problems confronting synchronic material constitution. These problems are interesting in this context, since if either Adams & Aizawa or (some) defenders of EC adopt the relation of ampliative constitution this carries with it its own set of metaphysical problems. Of the four problems that Wasserman focuses on, one problem in particular interests me. This is the so-called *grounding objection* raised against the plausibility of synchronic material constitution (for various ways of stating this objection, see Bennett 2004; Burke 1992; Simons 1987; and Zimmerman 1995). Let us focus again on the case of *David* and *Piece*. As we know, both *David* and *Piece* share the same matter. Also, the two objects share many of the same properties (e.g., weight, size, color, etc.). Commonly this aspect is taken to imply that *David* and *Piece* share many of the same categorical properties. Similarly, *David* and *Piece* also differ in many non-categorical properties such as conditions of existence (they are not identical with one another). But what could account for these differences? How can two things that are exactly alike in so many respects still differ in other respects?

Wasserman calls this the grounding objection of synchronic material constitution, "since it appeals to the common idea that non-categorical properties are grounded in categorical properties." (2009, p. 6) For instance, Baker (1999) attempts to explain this grounding in terms of *David* being a statue, and *Piece* being a lump of matter. Because *David* is admired as a piece of art, there are reviews written about *David*, and *David* exists as a statue relative to an art community. But, as Wasserman says, the "problem with this explanation is that it seems to get things exactly backwards, for it is natural to say that David is admired, reviewed, and discussed by those in the art community *because it is a statue* (rather than a mere lump of clay)." (2009, p. 7)

Another possible response would be to attempt to ground the non-categorical features of *David* and *Piece* in historical facts (Wasserman 2009, p. 7). But such a response – even if it might work in the case of *David* and *Piece* – will not work concerning EC. That is, references to causal-historical facts will not appeal to the EC theorist for the simple reason that such an appeal could too easily be utilized and altered to work as a critique of EC. In particular, Adams & Aizawa could say that it is metaphysically innocent to argue that relational properties can be accounted for by appeal to causal-historical facts. This they can infer from the externalist lessons on the individuation of mental content from Putnam (1975) and Burge (1979).

Hopefully it will now be clear why synchronic material constitution, regardless of the constitution relation being compositional or ampliative, is inconsistent with EC. In the next section, I attempt to contrast synchronic material constitution with diachronic constitution, while relating this to the C-C fallacy.

8.6. Synchronic material constitution, diachronic constitution, and the C-C fallacy

Lest the reader think that I am straying too far from matters of cognition, let me remind you that what is in dispute is the grain of fit between concepts in analytical metaphysics such as constitution and the nature of dynamical cognitive processes in EC. I have deliberately selected the most widely discussed examples in metaphysics – for instance, the relation between *David* and *Piece* – to establish what such an example tells us about the constitutive nature of objects, when that constitutive relation is supposed to hold synchronically. I have used this case contrastfully with cases of dynamical EC – e.g., the process of distributed remembering (Menary 2006) and van Gelder's (1995) case of the centrifugal governor.

One might wonder, of course, if the move from synchronic material constitution to what I term diachronic constitution is simply a bloodless coup? First, if my claim turns out to be correct, then this establishes that defenders of EC must avoid any blind adoption of the notion of constitution from metaphysics, because of the latter's incompatibility with common EC cases. Second, the need to pursue an alternative account of constitution points to something problematic with the concept of constitution as this notion is understood in metaphysics. In other words, if the relata we are investigating are inherently temporal (if they are *creatures of time*, as Noë would say (2006)) – which all relata are in dynamical systems – the tool-kit of material constitution cannot be applied to explain inherently temporal, dynamical phenomena. Dynamical systems, however, are ubiquitous in nature (Beer 2000; Friston & Stephan 2007; Kelso 1995; van Gelder 1998). Thus, the synchronic formulation of constitution as material constitution is applicable to only a small number of phenomena. Much more care and additional development of the notion of constitution is required to get at the nature of processes and other phenomena, where change in time and temporal unfolding is essential. In particular, since cognitive processes unfold in time continuous dynamical systems (Spivey 2007; Varela et al. 2001), we need to address these fine temporal details in order to identify the constitutive nature of just that which evolves over time.

In case the reader wonders whether the argument I am pursuing implies (a) restricting constitution to diachronic processes, and (b) to a relation between processes, this is not my intention. For example, I find it coherent to argue that constitution holds both (c) synchronically, and (d) between two or more distinct objects. The claim I find incoherent is the attempt to explain cases of (a) and (b) by applying the metaphysical tool-kit best suited to explain (c) and (d), because the relation of constitution that holds in dynamical systems such as cognitive systems is incompatible with how the relation of constitution is conceived of in metaphysics. Pushing the idea of diachronic constitution is meant to put in bold that we need to avoid exactly such a conflation and misapplication of phenomena and explanatory templates.

Before embarking on the task of explaining the idea of diachronic constitution, I wish here to map out a few interesting differences. The first of these differences is between constitution and causation. In her discussion of metaphysical relations, Bennett (2011) notes that causation as well as constitution is irreflexive, asymmetric, and transitive. If so, what differentiates causation, on the one hand, from constitution, on the other? Even if both causation and constitution are 'directed' in some relevant sense, these two dependence relations are typically understood to be wholly distinct. As I mentioned in chapters 2 and 3, the most commonly accepted additional feature of constitution that is not shared by causation is that for a relation to count as a relation of constitution it must hold synchronically, whereas causation is typically understood as a diachronic relation.

Although this is a tempting way to discriminate between causation and constitution, especially when considering dynamical cases such as distributed cognitive processes or modes of processing, on occasions temptation is best restricted. I think this is one of those occasions. Although she does not further develop this idea, Bennett herself stresses that we "should not require that building relations be *synchronic*." (2011, p. 94; italics added) As I mentioned, although she does not go on to develop this line of thought, it is important for my present purposes, in the sense that if some philosophers find my idea that metaphysical building relations can be diachronic obviously flawed (in some way), here we have a reputable philosopher of metaphysics stating that such an idea might not be so obviously mistaken. In fact, when conceivable as diachronically evolving, relations like constitution share far more features with certain modes of causation, especially what Clark (1997) and Wheeler (2005) call "continuous reciprocal causation" (CRC), than one might suspect.

We already have the idea that both causation and constitution can be diachronic, so I will leave this aside for now. What about the property of *asymmetry*? Even if it is standardly accepted that constitution and causation are asymmetric, in cases of EC, we should resist this assumption. Consider again Menary's process of remembering: "*X* is the manipulation of the notebook *reciprocally* coupled to *Y* – the brain processes – which together constitute *Z*, the process of remembering." (2006, p. 334; italics in original) This is the specific form of causation involved in CRC and nonlinear dynamics, in that, CRC "involves multiple simultaneous interactions and complex dynamic feedback loops, such that (i) each [process] partially determines, and is partially determined by, the causal contributions of larger numbers of other [processes], and, moreover, (ii) those contributions may change radically over time." (Wheeler 2005, p. 260)

What, then, about constitution? Craver & Bechtel note that all interesting cases of interlevel constitutive relations are *symmetrical* (2007, p. 553). Focusing on mechanisms, they state that: "The relation is symmetrical precisely because the mechanism as a whole is fully constituted by the organized activities of its parts; a change in the parts is manifest as a change in the mechanism as a whole, and a change in the [whole] is also a change in at least some of its component parts." (Craver & Bechtel 2007, p. 554) I will have more to say about the issues of temporality and symmetry concerning constitution in the section that follows.

The second difference that I wish to highlight is between how most EC theorists understand relations like constitution, on the one hand, and my account of diachronic constitution, on the other. For instance, Sutton (personal communication) thinks that constitution is synchronic. Sutton also thinks that constitution and causation are different. There is nothing problematic with the latter assumption, since what Sutton wants to claim is that if the disparate components are, in fact, part of a single cognitive system or process, then those components constitute that system or process. The causal interactions are not in themselves the ground for the constitution claim, though they are indeed a useful sign for the existence of such a distributed system or process. But this assumption is open to interpretation such that it is consistent with the one used in metaphysics: that there is a fundamental difference between causation and constitution, and that difference is that whereas the latter is synchronic, the former is diachronic. The problem, as I see it, is that this concedes too much to the critics. If there is a significant difference between causation and constitution, with that difference being that causation is diachronic (temporal) whereas constitution is synchronic (atemporal), then how does something that is inherently temporal (the complex causal relations between processes and their component parts at a lower level) atemporally constitute something that is inherently temporal at a higher level (e.g., the distributed process of remembering)? In the end, then, Sutton's view is open to the same kind of worry that I raised previously: how can the relation of constitution that holds between processes be atemporal (i.e., synchronic) if the relata themselves (at both higher- and lower-levels) are essentially temporal (i.e., diachronic)?

This brings me to the final difference I want to focus on in this section, namely that between the C-C fallacy and diachronic constitution. Specifically, if a defender of EC were to work with diachronic constitution, this defender would not commit the C-C fallacy. First, the C-C fallacy turns on an argument that conflates intralevel causal relations with interlevel constitutive relations, in the sense that the C-C fallacy works only if an inference is made from causation to constitution, where these relations are said to persist on the same level, e.g., that of the constituents. Second, the C-C fallacy assumes that constitution is itself synchronic – this should be evident since the C-C fallacy turns on compositional constitution. But since cases of distributed cognitive processes are temporal, and synchronic notions of constitution fail to pick out the fine temporal details essential to what it is to be a process, the C-C fallacy is question begging.

8.7. Diachronic constitution

To further unravel the notion of diachronic constitution, I will discuss several core features of diachronic constitution in turn in this section.

8.7.1. "Small-m" mereology, not "big-M" mereology

It is quite intuitive to associate part-whole relations with relations of constitution, and because appeals to the formal ideas of the theory of extensional mereology has been quite influential in metaphysics, perhaps, then, we should also think of diachronic constitution as consistent with the formal part-whole theory of extensional mereology? Burrowing a distinction from Wilson (who modifies this distinction from Simons (1987)), I now show why diachronic constitution can (and should) be expressed without any appeal to extensional mereology. The relevant distinction is that between *small-m* mereology and *big-M* mereology, with the latter referring to the specific formal theories of Mereology that grew out of Lesniewski's *Foundations of a General Theory of Manifolds* (1916) and Leonard & Goodman's *The Calculus of Individuals* (1940). The primary concern with the notion of big-M mereology is that it construes the part-whole relation as a partial ordering, *viz*., as an antisymmetric, reflexive, and transitive relation. Recall that in section 8.5.1, I argued that regardless of considering the relation between *David* and *Piece*, or the relation between *Z* (process of remembering) and *X* (manipulation of notebook) and *Y* (brain process), neither *David* nor *Piece* and neither *Z* nor *X* & *Y* can plausibly be thought of as part(s) of itself (themselves). Perhaps even more problematic, the extensionality principle of big-M mereology violates the commonly accepted idea that constitution is a relation of non-identity. Given the attention from constitution theorists on distinguishing the relation of constitution from a relation of identity, I follow Wilson's advice in thinking that "it would seem prudent to avoid building this into one's view of constitution from the outset." (2007, p. 7) This is, of course, still fully consistent with conceiving of small-m mereology as in line with diachronic constitution.

8.7.2. Process ontology 'yes', but non-eliminative

While this chapter – and my account of diachronic constitution – is not intended as a defense of process ontology and does not offer a comparison of such views with alternative ontological models, such as traditional substance ontology, and various competitor views, e.g., Whitehead's event ontology, trope ontology, stage ontology, and so on – tasks for another occasion – diachronic constitution shares a *kinship* with some form of noneliminative process ontology. First, unlike certain eliminative variations of process ontology such as French & Ladyman's (2003) account of ontic structural realism, who argue that out best physics is incompatible with ontological categories such as "individuals," on my account of diachronic constitution, the very idea of processes presupposes that processes have *individual parts*. A process might involve any number of component parts, but it always involves *some* parts. Even though processes themselves may occupy the relata in relations of constitution (as in the relation between *Z*, the process of remembering, and *X* & *Y*, the process of manipulating a notebook reciprocally coupled to brain processes), we need to be aware that our analysis, and subsequently explanation, must stop somewhere. That is, even if our best physics tells us that individual entities do not exist – e.g., only quantum fields exist – this will not make much sense in the context of cognitive science. In cognitive science we want to be able to locate and preferably identify entities as well as their activities (Bechtel & Richardson 1993). Note, though, that there is an important and non-trivial difference between how process ontologists (see e.g., Seibt 2009) conceive of individual parts and how the notion of "part" is preserved and propagated in accounts of synchronic material constitution.

To appropriately characterize "part" in processual terms, we must replace what Seibt has recently called the "particularist conception of individuals" – i.e., entities that are intrinsically individuated and which have a determinate unique location – with a view of individual parts that focuses not so much on "location but on '*specificity-in-functioning*' in the widest sense of 'functioning,' i.e., focuses on the dynamic role of an entity (e.g., an activity) *within a certain dynamic context*." (2009, p. 484; italics added). This is a crucial difference between accounts of synchronic material constitution and the idea of diachronic constitution for a couple of reasons. First, both compositional constitution and ampliative constitution presuppose that the physical constituents are intrinsically individuated – this we can see in the intrinsic necessitation constraint and in the relational/intrinsic constraint. If processes, according to process ontologists, are individuation-dependent upon the larger context within which they are embedded for their dynamical function, processes cannot be intrinsically individuated (for similar points, see chapters 6 and 7). Second, moving from a particularist notion of individuals to a view of component parts as individuated qua their specificity-in-functioning is indicative of a shift away from focusing on the material as well as spatial co-*location* of relata in constitutive relations to a practice of individuating aspects of nature in terms of *dynamic function* – *viz*., in terms of what is happening or is going-on in situated context.

Framing the constitutive thesis of EC in terms of ontological frameworks akin to noneliminative modes of process ontology is consistent with particular strands of EC theorizing. For instance, Menary (2012) distinguishes between "artifact extension" (AE) and "enculturated cognition" (EnC). AE is the version of EC advocated by Clark & Chalmers (1998), Clark (2008), and Wheeler (2010). EnC, on the other hand, appeals to the idea that cognitive processes are driven and partly constituted within a species of cultural practices (Hutchins 2008, 2011a; Menary 2007, 2009), and motivates a shift away from a focus on "things" to an enactivist approach to cognition as the unfolding of dynamical processes and/or patterns (see also Chemero 2009; Di Paolo 2009; Hutchins 2011b; Varela et al. 1991). As far I can tell, EnC is process-based. It is in virtue of this that cognitive processes involve multiple feedback loops and organizing activity across the boundary of the organism itself, which reveals incoherence in the notion of "intrinsically individuated constituents".

To proceed further with my analysis of why diachronic constitution shares a kinship with specific strands of noneliminative process ontology, what we need, to get a firmer grip on these issues, is an example to analyze. Let us consider, then, the dynamical system I foregrounded in section 8.2, the Watt (centrifugal) governor. It is not too surprising, I think, that dynamical systems (and distributed processes such as processes of remembering) do not dovetail with explanations in terms of synchronic material constitution, because analytical metaphysics is not well equipped to deal with dynamic phenomena in general. The insights from dynamics throw into question how we should understand the notion of "part," if indeed we keep (as I have argued) the requirement that processes involve component parts as an element of our ontology of processes. Unlike synchronic views of material constitution, where constituent parts are *particulars* (i.e., "entities that (i) each have a determinate *unique* spacetime location and (ii) have this location necessarily since they are individuated in terms of [their] location." (Seibt 2009, p. 484)), processes are best understood as having *non-particular* component parts.

To appropriately describe parts in processes, we must move away from what Seibt calls the "particularist conception of individuals," and replace it with a perspective that puts emphasis "not on location but on '*specificity-in-functioning*' […], i.e., focuses on the dynamic role of an entity (e.g., an activity) *within a certain dynamic context*." (2009, p. 484; italics added) Applying Seibt's (2009) model of processes to the Watt governor, we can say the following. First, processes are *temporally extended*, i.e., there is no such thing as an instantaneous process. In the case of the Watt governor, the process of maintaining a constant speed of the flywheel does not take place at a time instant *t*; rather, it is the unfolding of a complex pattern or process over time. Second, processes do not necessarily occur in a unique spatiotemporal location – ontologically speaking, a process is *not a particular*. Albeit mechanically organized, which of course limits the freedom of movement in the Watt governor, the throttle valve, the arm angle, the spindles, the pulley belted to the flywheel, and the collar slides are all in continuous and mutually influencing interaction. Third, processes must be *individuated in terms of their roles in a dynamic context* – that is, because they are non-particulars, they must be individuated so. In dynamic systems theoretic terms, we can explain the relationship between the steam engine and the governor system such that the arm angle of the governor, call this θ, is a parameter of the engine system, whereas the engine speed, call this w, is a parameter of the governor system (van Gelder 1995, p. 357). This relationship between θ and w is known as *coupling*, which enables us to explain the dynamical behavior of θ and w as comprising what van Gelder refers to as a "single dynamical system in which both arm angle and engine speed are state variables." (1995, p. 357)

Similarly to θ and w, we can think of the relationship between an agent (a brain, perhaps) and its environment as two dynamical systems *A* and *E* dynamically coupled to one another, and where both *A* and *E* are time-continuous dynamical systems. Beer represents this coupling as follows: *S* is a sensory function from environmental states to agent parameters, and *M* is a motor function from agent state variables to environmental parameters, with $S(X_E)$ standing in for an agent's sensory inputs, and $M(X_A)$ corresponding to its motor outputs. As Beer shows, this gives us the following equations: $X_A = A(X_A; S(X_E); U'_A)$, and $X_E = E(X_E; M(X_A); U'_E)$ (1995, p. 181). Given the continuous reciprocal coupling between *A* and *E*, Beer emphasizes that we can see – just like the case with θ and w – "the two coupled nonautonumous systems *A* and *E* as a single autonomous dynamical system *U* whose state variables are the union of the state variables of *A* and *E* […]." (1995, pp. 182-83).

How does this relate to the C-C fallacy? First, if a process *x* occurs in *y* and *y* is causally interacting with *z*, it does not follow (so Adams & Aizawa argue) that *x* "extends into" *z*. This is the form of Adams & Aizawa's example of the air conditioning system: if evaporation occurs in an evaporation coil and the latter is causally linked to a compressor coil, it does not follow that evaporation "extends into" the compressor coil. Nobody, I believe, would dispute this. But, as I argued, this template is deeply problematic for the simple reason that most defenders of EC do not argue in a way corresponding to that template. Indeed, having made the distinction between a "particularist conception of individuals" and a "non-particularist conception of individuals," it is easy to show that Adams & Aizawa are indeed committed to the former conception of individual entities as *particulars*. This is a serious point, since it is an incoherent assumption when applied to processes as relata in relations of constitution. Consider that "something is a particular if by necessity it occurs in one entity only." (Seibt 2010, p. 29) In other words, *x* occurs in *y* if by necessity it only occurs in *y* – evaporation occurs in *y* if by necessity it only occurs in *y*. But, as we have seen, processes cannot be explained by reference to such necessitation and location-exclusive requirements. Thus, working with particularist assumptions presents a problem for Adams & Aizawa.

8.7.3. Counterfactual dependency

To further highlight some of the similarities between causation and a diachronic notion of constitution, together with providing an answer to a possible objection from Adams & Aizawa, I will now consider the issue of counterfactual dependency.

There are many ways by which to attempt to discriminate between causal dependency and constitutive dependency. One way, presented in this chapter, is to assume that only causation is temporal, whereas constitution is atemporal. Of course, this may hold only in cases where we contrast synchronic material constitution with causation – it does not hold, I have argued, once we contrast a diachronic and process-based notion of constitution with causation, since both of these are temporal. Another possibility seems to be to explain causal dependency in terms of counterfactual dependency, and from this try to show that only causal dependency can be explained counterfactually, whereas counterfactual dependency is insufficient to justify constitutive dependency. This, I suspect, is yet another assumption that is driving Adams & Aizawa's insistence that defenders of EC commit the C-C fallacy.

In this section my aim is to establish that diachronic constitution is immune to such an accusation. I shall show that one cannot appeal to counterfactual dependency in order to discriminate between causation and constitution.

The basic idea of analyzing causation in terms of counterfactuals is that causal claims can be made understandable as well as explained in terms of counterfactual conditionals of the form: if *C* had not occurred, *E* would not have occurred. But why think that causation (or causal claims) is conceptually linked with counterfactuals? According to Menzies, one "reason is that the idea of a cause is conceptually linked with the idea of something that makes a difference and this idea in turn is best understood in terms of counterfactuals." (2008, p. 4) Or, in the words of Lewis: "We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its effects – some of them, at least, and usually all – would have been absent as well." (1973, p. 161)

Consider, firstly, the following schematic claim: a cognitive process *Z* causally depends on two other processes, *X* and *Y*, just in case if *X* and *Y* had not occurred *Z* would not have occurred. Consider, secondly, our familiar example: this process of remembering is constituted by processes of manipulating a notebook jointly and reciprocally coupled to brain processes. This statement can be causally explained in terms of counterfactual conditionals of the following form: *Z* causally depends on *X* and *Y* just in case if *X* and *Y* had not occurred *Z* would not have occurred. Insofar as the counterfactual conditional "if *X* and *Y* had not occurred, *Z* would not have occurred" entails the causal statement "*X* and *Y* cause *Z*," it seems that Adams & Aizawa could argue: given such an entailment of the counterfactual conditional, the defender of EC is still committing an instance of the C-C fallacy, since it is not enough to show that *Z* is causally dependent on *X* and *Y* if the target is to establish that *Z* is constitutively dependent on *X* and *Y*.

However, for Adams & Aizawa to justifiably make this claim, they need to establish the additional claim that the dependency expressed by counterfactuals is limited to causal dependency. But such a claim they will be unlikely to construct successfully. In his (1973), "Causes and counterfactuals," Kim points out that the "sort of dependency expressed by counterfactuals is considerably broader than strictly causal dependency and that causal dependency is only one among the heterogeneous group of dependency relationships that can be expressed by counterfactuals." (1973, p. 570) There are cases involving processes, whose persistence is dependent on spatiotemporal continuity, and in which one event is a constituent part of another. Consider, for example, the case from chapter 3, namely Hofweber & Velleman's example of the process of writing a cheque:

> "A process of writing a cheque is a temporally extended process, with temporal parts consisting in the laying down of each successive drop of ink. What there is of this process at a particular moment – the laying down of a particular drop – is not sufficient to determine that a cheque is being written, and so it is not sufficient to determine which particular process is taking place. […] Not only, then, is the process not present in its temporal entirety within the confines of the moment: it is not fully determined by the events of the moment to be the process that it is." (2011, p. 50)

In this case, my laying down of each successive ink drop, I_1, …, I_n, is a constituent event in the overall process of writing a cheque; and, following Kim, it is probably true to say: 'If I had not laid down each successive ink drop, I_1, …, I_n, I would not have written a cheque'. But, it is unlikely that my putting an ink drop down followed by another drop of ink causally determines me writing a cheque. Hence, the first key point is: "counterfactual dependency is too broad to pin down causal dependency." (Kim 1973, p. 571)

If this is the case, then Adams & Aizawa cannot straightforwardly identify the case of distributed processes of remembering with the C-C fallacy. Recall: "X is the manipulation of the notebook *reciprocally* coupled to *Y* – the brain process – which together constitute *Z*." (Menary 2006, p. 334; italics in original) This statement can be given both a causal and a constitutive explanation, and both of these can be analyzed in terms of counterfactuals. In contrast to '*Z* causally depends on *X* and *Y* just in case if *X* and *Y* had not occurred *Z* would not have occurred,' this statement can also be understood as saying '*Z* constitutively depends on *X* and *Y* just in case if *X* and *Y* had not occurred *Z* would not have occurred. Therefore, Adams & Aizawa cannot use counterfactual analysis to pin down a distinction between causal dependence and constitutive dependence.

8.7.4. Hybridity

Suppose Adams & Aizawa accept the claim that counterfactual dependency cannot motivate (underpin) the C-C fallacy. Nevertheless, in the context of my claim that distributed cognitive phenomena are hybrid, consisting (among other things) of both causal and constitutive relations, Adams & Aizawa could still argue that if diachronic constitution does not rely on "inferring constitution from causation (or coupling), then *that* is not a defense of what other extended cognition theorists have said." (Adams & Aizawa 2008, p. 104; italics in original) Rather, it "seems to be an abandonment of the coupling [causation] to constitution arguments, […]." (Adams & Aizawa 2008, p. 104) First, they would be absolutely correct to specify that insofar as the defenders of EC adopt the account of diachronic constitution, they are not making an inference from causation to constitution. This is because causal relations are horizontal relations, whereas the constitution relation, at least when understood to hold diachronically, is neither horizontal nor vertical but rather diagonal. Second, Adams & Aizawa might also be correct to point out that such an account might not qualify as a proper defense of what other (previous) EC theorists have said. But why is that a problem? Indeed, the C-C fallacy itself turns on there being a fallacious inference from causation to constitution. If there is no such fallacious inference on my account, this is *not* a problem for me; rather, it is a problem for Adams & Aizawa. In particular, it seems that Adams & Aizawa have a problem with any account that abandons such an inference, since such an account would overcome (and therefore be immune to) the accusation of unjustifiably inferring from facts about causation to facts about constitution.

8.7.5. Top-down and bottom-up constitutively mediated effects

Many assumptions about causation and constitution turn on the idea that both of these relations hold asymmetrically. However, even though this might imply that causes precede their effects (that is, there is an asymmetric relation from a cause to an effect), it does not follow that an effect could not have feedback effects on its cause. This is precisely the view of causation that Clark (1997) and Wheeler (2005) refer to as CRC. The same can be said about constitution. In distributed remembering, *X* and *Y* constitute *Z* asymmetrically, in the sense that *X* and *Y* constitute *Z* and *Z* does not constitute *X* and *Y*. However, once constituted, *Z* may display top-down effects on its constituents (a point I made much of in chapter 7).

What I hope to have shown so far is that if Adams & Aizawa argue that defenders of EC confuse constitution with causation, it is in fact Adams & Aizawa that violate the central idea that causation is an intralevel phenomenon, whereas constitution is an interlevel phenomenon. In their discussion of why it is erroneous to suppose that causation works across different levels, Craver & Bechtel (2007) use an example from Patricia Churchland (1993), who expresses a similar worry concerning causation as an interlevel relation. The example of choice is from the *Betty Crocker Cookbook*. As Craver & Bechtel re-iterate Churchland's claim:

> "Betty correctly explains that microwaves work by accelerating the component molecules in the food. However, she takes a decidedly wrong turn when she explains further that the excited molecules rub against one another and *generate* heat through fiction. Betty's error, of course, is in supposing that

> heat is causally produced by the increase in mean kinetic energy when in fact heat is *constituted* by their mean kinetic energy. The causal reading in this case is simply erroneous." (2007, p. 555)

Similarly with our case of *Z*, the process of distributed remembering. The flaw that Adams & Aizawa commit is similar to Betty's error, in that, they charge the defenders of EC with the claim that *X* (the process of manipulating the notebook) reciprocally coupled to *Y* (brain process) is what *generates* (i.e., causes) *Z*, when, in fact, *Z* is constituted by *X* and *Y*.

How can we express the idea that constitution, as an interlevel relation, is symmetric? Enter the second non-cognitive example, the classical example from dynamical systems theory of a fluid heated from below and cooled from above. In the previous chapter, we looked at this particular example in detail; hence, here I briefly describe the example again, and pick out what is of important given the subject matter being discussed here.

The phenomenon in question is *convection rolls*. Of course, I do not want to claim that EC simply is like a fluid composed of homogenous elements. Far from it, since most cases of EC consists of a multiplicity of heterogeneous elements. But this dynamical, non-cognitive, example highlights in an easy to understand way what the dynamicists refer to as a *collective variable*, which is the kind of mechanism that will allow me to show just how constitution can be symmetric. Here is how the example goes. Take some oil, put it in a pan, and apply a heat source from below. As the heat is applied it increases the difference in the temperature between the top and the bottom of the oil layer. At a critical threshold, an event called "an instability" occurs such that the liquid begins to self-organize a coherently rolling motion. This motion is a convection roll. What happens is that the cooler liquid at the top is denser (and heavier), thereby falling, whereas the liquid at the bottom is warmer (therefore less dense and lighter), thus tending to rise to the top. The resulting convection roll is called a collective or cooperative effect in the language of dynamical systems theory. The temperature gradient itself is referred to as a control parameter, yet it is not a parameter that encodes or pre-specifies the pattern of convection rolls. What is fascinating here is that a pan of oil may contain something on the order of 10^{20} molecules (Kelso 1995, p. 8) all subject to random disordered motion. However, once the rolling motion begins, the convection rolls ensure that "all parts of the liquid no longer behave independently but are sucked into an ordered, coordinated pattern that can be described precisely using the parameter concept [*viz*., by using the order parameter or collective variable concept]." (Kelso 1995, p. 8) The collective variable (*viz*., the rolling motion of convection rolls) is *constituted* by the collective cooperation of the individual parts of this dynamical system, here the fluid molecules. Yet the collective variable "governs or constrains the behavior of the individual parts." (Kelso 1995, p. 8) On the one hand, the component parts constitute the whole, yet the whole can affect the behavior of its parts, on the other. On my view, then, the interlevel relationship is constitutive, and because the constitutive effects run from both bottom-up and from the top and down, this is why diachronic constitution is symmetric. Because of this, by applying the distinction between constitution as an interlevel relation and causation as an intralevel relation, we can identify the interlevel constitutive relation as between the convection roll (or rolls) and its component parts without having to fallaciously appeal to this relationship as a form of causation.

Such mediated top-down and bottom-up effects between diachronically unfolding processes can be usefully applied to cases of distributed cognitive processes. Hence, I now turn to the third, and this time cognitive, example.

Here I consider the example of transactive memory that I dealt with in chapter 5. Recall, Wegner (1987) introduced the concept of "transactive memory systems" (TMSs) in an attempt to explain how individual members in long-tenured groups, intimate couples, and so on, rely on each other to obtain, process, and communicate knowledge from different domains. As Harris et al. state, remembering "often occurs jointly in social groups" (2011, p. 268; see also Barnier, Sutton, Harris & Wilson 2008). As Harris et al. go on to say: "People in close relationship are likely to be behaviorally, emotionally, and cognitively 'interdependent' […] – that is, in collectives such as couples, families, […], and work teams, remembering is an interactive activity where memories are dynamically and jointly constructed […]." (2011, p. 268)

Often, yet not always, such dynamically and collaboratively constructed modes of remembering result in both a division of labor and a specialization of knowledge between couples, friends, work teams, etc. Lewis puts this nicely, when she says: "According to transactive memory theory, group members divide the cognitive labor for their tasks, with members specializing in different domains. Members rely on one another to be responsible for specific expertise such that collectively they possess all of the information needed for their tasks." (2003, p. 587) Transactive memory theory describes both the processes involved in actual instances of transactive memory and the benefits for memory that may occur when remembering is shared between two or more individuals (see e.g., Barnier et al. 2008; Harris et al. 2011; Lewis 2003; Theiner & O'Connor 2010; Wegner 1987). A TMS, then, is a cooperative and mnemonic division of labor in learning, remembering, and communicating within dyads, triads or larger social groups.

Where there is a mnemonic division of labor, the differentiation and socially distributed processes of retrieval, encoding, and sharing of autobiographical memory, may result in an integrative process of socially distributed remembering. Furthermore, collective remembering – just like convection rolls, the Mexican wave, the workings of the centrifugal governor, and so on – depends for its existence on spatiotemporal continuity. For transactive remembering to persist it must persist as a dynamical unfolding in real time. Of course, even in dynamical cases such as transactive remembering one might insist that despite the process of transactive remembering being time continuous this does not prevent one from describing or explaining such a phenomenon synchronically.

For instance, dynamical systems theory conceptualizes systems geometrically, that is, in terms of regions in a *state space* (i.e., distances and trajectories in a space of possible states). A cognitive scientist, for example, may use the mathematical paradigm of dynamical systems theory to point to the position of a system in a dynamical state space, locating the system at *T2* over an interval *T1*, …, *Tn*. This is an example of what Spivey calls a "kind of coarse grained averaging measurement" (2007, p. 30), in the sense that synchronic explanations – metaphysically understood – can at best be an abstraction or a product of a particular model enabling a scientist to describe or explain that phenomenon as existing somewhere on a dynamical trajectory at a particular clock time *T2*. But we should not mistake this as genuine evidence for the system actually resting in a discrete state at *T2*. As. Spivey says: "[Claiming] that a system was in a particular "state," **X**, at a particular point in time, really boils down to saying that the *average* of the system's states during that *period of time* was **X**. This kind of coarse averaging measurement is often a practical necessity in science, but should not be mistaken as genuine evidence for the system actually resting in a discrete stable state." (2007, p. 30; italics in original)

According to Barnier et al. (2008, p. 38), transactive memory theory predicts that recalled memories by individuals in diachronically unfolding retrieval processes would be more than the sum of individual memory. That is, the constituted process of successful transactive remembering should have emergent properties that are not only greater than, but also *different* from the sum of individual memory – either in quantity (amount of information remembered) or in terms of quality (e.g., the emotional richness of the particular jointly remembered event) – see Chapter 5 for more detail on these diachronically emergent properties.

Once the process of transactive remembering is initiated between, for example, two individuals, they will begin to collaborate in a coherent fashion. The TMS is no longer merely two separate individuals with their individual memories. Instead, they cooperate to create a coherent and dynamically shared pattern of collaborative memory unfolding over time. This *new* version will affect, in a top-down fashion, their individual memories of the event, and it is "quantitatively and qualitatively different from what each remembered alone." (Barnier et al. 2008, p. 38) In the language of dynamical systems theory, we can say that the parts (the individuals, for example) no longer behave independently but "are sucked into an ordered, coordinated pattern […]." (Kelso 1995, p. 8), and this pattern has top-down constitutively mediated effects on the component parts. It is important to mention here that such top-down effects are operative in both *occurrent* instances of transactive remembering – such as the dialogue between the husband and the wife surveyed in chapter 5 – and in long-term, nonoccurrent modes of individual's transactive memories. Consequently, in transactive memory, we come across top-down constitutively mediated effects in the here-and-now and across developmental time.

8.8. Conclusion

The empirical and theoretical approach I have developed in this chapter departs from much of the core and mainstream literature in both metaphysics and EC. I have argued that if we wish to understand and explain the constitutive relation in cases of distributed cognitive processes, we cannot rely on the traditional framework of material constitution in metaphysics. Instead, what is needed is a notion of constitution that shares a kinship with noneliminative process ontology and that is inherently diachronic. I have argued that this notion of diachronic constitution breaks with how constitution is typically considered in the debate about EC. In relation to this debate, I have attempted to show that the critics – Adams & Aizawa – are wrong to criticize defenders of EC with committing the C-C fallacy, because Adams & Aizawa are working with a notion of synchronic compositional constitution that is inconsistent with common cases of EC.

9. Conclusion

I began this thesis by setting out two research aims. The first aim, in particular, concerned the extended cognition thesis (EC), where I advanced the claim that it is possible to radicalize EC further than it is already considered to be. Indeed, despite EC's groundbreaking implications for *where* cognitive processes and processing may be instantiated or exemplified – expanding the traditional view of the brain as the locus of cognition to include parts of the non-neural body and local environment – I put forth the claim that there is a tendency within the EC literature to adopt – without further scrutiny – key notions from metaphysics such as composition, constitution, supervenience, realization, and emergence. With respect to this adoption of metaphysical concepts, I have attempted to show throughout the thesis that these concepts of metaphysical building relations, insofar as they are simply redeployed in EC without significant reformulation, carry with them conceptual baggage unsuited to analyze the *relata* and *systemic dynamics* in many examples of EC.

As a result, and in relation to the first research aim, the thesis has been an attempt to establish the need for a reformulation of the metaphysical foundation of EC by (a) pointing to inconsistencies between the received synchronic view of metaphysical relations of dependence, on the one hand, and dynamical cognitive systems and processes, on the other, and (b) by developing an alternative diachronic account.

Overall I have argued that the DIACHRONIC perspective developed in the thesis finds a congenial environment in contemporary dynamical systems theoretic approaches to cognition, including second-wave versions of EC (Menary 2007, 2010a; Sutton 2008, 2010), certain strands of enactivism (Di Paolo 2009; Engel 2010; Varela et al. 1991) and dynamical cognitive science (Chemero 2009; Spivey 2007; van Gelder 1995, 1998). Common for all of these accounts of cognition is that they start from the premise that cognitive processes are embedded within and instantiated by self-organizing, nonlinear, and temporal activity in dynamical systems, whose interdependent components may include, in the right circumstances, neural, bodily, and local environmental elements.

On these accounts, then, and supported by DIACHRONIC, every cognitive function, from perceptual-motor behavior to advanced forms of problem-solving and kinds of remembering, arise through concurrent integration and transformation of functionally distinct as well as topographically different neural regions, together with regions distributed across neural, bodily, and worldly resources. With respect to this, the central issues occupying me throughout the thesis – especially concerning how to think about metaphysical building relations – has been the following: that the integration of these disparate regions occur over multiple different timescales of duration, ranging from the very short (milliseconds, seconds) to the much longer (minutes, hours). Insofar as both the relata in dynamically distributed cognition and the mechanisms through which the relata jointly combine to produce cognitive outcomes are inherently temporal, the question that has motivated the development of DIACHRONIC was: how can a metaphysical building relation – e.g., constitution or composition – that holds between processes be atemporal? That is, if the very nature of a process is to unfold across a region of spacetime for a process to be what it is, then how can the existence of a process be determined at an atemporal instant? Throughout the thesis, I argued that it is this particular dilemma that defenders of EC face insofar as they keep intact the received synchronic view of metaphysical relations of dependence.

With so much evidence in support for the view that cognitive processes are continuously dynamical and richly temporal, crisscrossing levels and boundaries, there is no principled reason for why we should keep endorsing the static metaphysical framework expressed by received views of constitution, composition, emergence, supervenience, and realization. Indeed, the central claim of the thesis has been that if cognitive systems and cognitive processes are ineliminably temporal, then any robust metaphysics of such cognitive systems and processes must reflect such inherent diachronic characteristics.

By developing a diachronic metaphysical framework, it has been my aim to move the EC research forward into a *third-wave of EC theorizing*, where a clear conceptual framework for a dynamical ontology underpins genuinely dynamical relations. That is what the project of reformulating the metaphysics of EC has been all about.

Despite its congeniality with dynamical systems theoretic approaches to cognition, the proposed DIACHRONIC account, however, finds significantly less congeniality in analytical metaphysics insofar as the received view of metaphysical building relations (in this area of philosophical inquiry) is most commonly expressed in terms of SYNCHRONIC. This was what motivated the second aim of the thesis, namely to broaden the boundaries of metaphysical theorizing about dependence relations to not only include SYNCHRONIC but equally DIACHRONIC. What I have not done is to argue that the standard synchronic view in metaphysics is false *simpliciter*; rather, my ambition has been to show that when certain conditions are operative – for example, when both relata in question are processes, and because processes require for their persistence spatiotemporal continuity – the received synchronic account of metaphysical building relations is ill equipped to analyze the specific relations of dependence involved. Note, though, that insofar as all physical systems are dynamical systems, this carries with it the implication that the synchronically motivated view of metaphysical dependence relations does not apply to physical systems *qua* dynamical systems. That is a controversial claim, especially because of the inferences that I have drawn from it in this thesis. For example, in chapter 6, I argued that when considering the relationship between the free energy minimization and the realization relation, it is far from certain that the realization is the appropriate relation by which to understand the relationship between free energy minimization and the mechanisms instantiating free energy minimization. Also, in chapter 7, where I argued that the standard view of emergence as a supervenience relation is problematic insofar as our explanatory target is to understand how emergent properties arise in dynamical systems.

In spite of these implications for SYNCHRONIC, I have not shown that DIACHRONIC is the only true account about metaphysical dependence relations insofar as dynamical systems are concerned. I do not think that DIACHRONIC is the only true account for the simple reason that it is has not been possible to address all the relevant issues in the thesis. For example, there is still much work to be done in terms of understanding the parthood relation in dynamical systems. Also, we need an account of time. That is, if, as I argued in chapter 4, the notion of an ontological synchronic conception of time is problematic, since even the *here-and-now* involves a temporally rich unfolding of activity, this raises the question of how we should understand synchronic statements in cognitive science, since most cognitive scientists would insist that it is correct to describe a system as being in a certain state, say, at a certain point in time. Furthermore, insofar as DIACHRONIC turns out to be correct, we still require an explanation of how DIACHRONIC is different from causation given the fact that DIACHRONIC is not a relation of causation. In other words, if DIACHRONIC is a non-causal relation of dependence, yet a temporal relation, how, then, does DIACHRONIC differ from the relation of causation? These are just a few questions that I think needs to be settled before we can say that DIACHRONIC is *the* account in town – issues to be addressed in another project.

Nevertheless, what I hope to have done is to have made a convincing case for the need to move away from the received synchronic view of metaphysical dependence relations and towards a diachronically infused alternative. Furthermore, that doing so will prove fruitful for an understanding of metaphysical relations of dependence in dynamical cognitive systems and dynamical cognitive processes.

Michael D. Kirchhoff
Sydney
December 2013

References

Adams, F., & Aizawa, K. (2008). *The Bounds of Cognition*. New York: Blackwell.

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, 14(1), 43-64.

Aizawa, K. (2013). Introduction to "the material bases of cognition". *Minds & Machines*. DOI: 10.1007/s11023-013-9312-8.

Aizawa, K. (2010). The coupling-constitution fallacy revisited. *Cognitive Systems Research*, 11, 332-342.

Aizawa, K., & Gillett, C. (2009a). The (multiple) realization of psychological and other properties in the sciences. *Mind & Language*, 24(2), 181-208.

Aizawa, K., & Gillett, C. (2009b). Levels, individual variation and massive multiple realization in neurobiology. In Bickle, J. (ed.), *Oxford Handbook of Philosophy and Neuroscience* (pp. 539-582). Oxford: Oxford University Press.

Ashby, R. (1952). *Design for a Brain*. Chapman & Hall.

Baker, L. B. (2000). *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.

Baker, L. B. (1999). Unity without identity: A new look at material constitution. *Midwest Studies in Philosophy*, XXIII, 144-165.

Baker, L. B. (1997). Why constitution is not identity. *Journal of Philosophy*, 94, 599-621.

Ballard, D., Hayhoe, M., Pook, P., & Rao, R. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 7(1), 66-80.

Barnier, A. J., Sutton, J., Harris, C. B., & Wilson, R. A. (2008). A conceptual and empirical framework for the social distribution of memory: The case of memory. *Cognitive Systems Research*, 9, 33-51.

Baxter, D. (1988). Identity in the loose and popular sense. *Mind*, 97, 623-635.

Beach, K. (1988). The role of external mnemonic symbols in acquiring an occupation. In M. M. Gruneberg and R. N. Sykes (eds.), *Practical Aspects of Memory* (pp. 342-346). New York: Wiley.

Bechtel, W. (2009). Molecules, systems, and behavior: Another view on memory consolidation. In J. Bickle (ed.), *The Oxford Handbook of Philosophy and Neuroscience* (pp. 13-40). Oxford: Oxford University Press.

Bechtel, W. (2008). *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. New York & London: Routledge.

Bechtel, W., & Richardson, R. C. (1993). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, NJ: Princeton University Press.

Beer, R. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Science*, 4(3), 91-99.

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

Bedau, M. (1997). Weak emergence. *Philosophical Perspectives*, 11, 375-399.

Belyaev, V. B., Motovilov, A. K., Sermyagin, A. V., Kuznetzov, I. V., Sobolev, Yu. G., Smolnikov, A., Klimenko, A. A., Osetrov, S. B., and Vasilev, S. I. (2001). Search for nuclear reactions in water molecules. *Physics Letters B*, 552(3-4), 222-226.

Bennett, K. (forthcoming). *Making Things Up*. Under contract with Oxford University Press.

Bennett, K. (2011). Construction area (no hard hat required). *Philosophical Studies*, 154, 79-104.

Bennett, K. (2004). Global supervenience and dependence. *Philosophy and Phenomenological Research*, 68(3), 501-529.

Bickhard, M. (2004). Process and emergence: Normative function and representation. *Axiomathes*, 14(3), 135-169.

Block, N. (1997). On a confusion about a function of consciousness. In N. Block, O. Flanagan, and G. Guzeldere (eds.), *The Nature of Consciousness* (pp. 375-416), Cambridge, MA: The MIT Press.

Block, N. (1980). Introduction: What is functionalism? In N. Block (ed.), *Reading in Philosophy of Psychology* (Vol. 1, pp. 171-184). Cambridge, MA: Harvard University Press.

Boogerd, F. C., Bruggeman, F. J., Richardson, A., Stephan, A., & Westerhoff, H. (2005). Emergence and its place in nature: A case study of biochemical networks. *Synthese*, 145, 131-164.

Broad, C. D. (1925). *The Mind and its Place in Nature*. London: Routledge & Paul Kegan.

Brogaard, B. (2000). Presentist four-dimensionalism. *The Monist*, 83(3), 341-356.

Brooks, R. (1999). *Cambrian Intelligence*. Cambridge, MA: The MIT Press.

Brown, E. C., & Brüne, M. (2012). The role of prediction in social neuroscience. *Frontiers in Human Neuroscience*, 6, 1-19.

Brown, H., Friston, K., & Bestmann, S. (2011). Active inference, attention, and motor preparation. *Frontiers in Psychology*, 2(1), 1-10.

Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy*, 4, 73-122.

Burke, M. (1992). Copper statues and pieces of copper: A challenge to the standard account. *Analysis*, 52, 12-17.

Campbell, R. J., & Bickhard, M. (2011). Physicalism, emergence, and downward causation. *Axiomathes*, 21, 33-56.

Cash, M. (2013). Cognition wihout borders: "Third-Wave" socially distributed cognition and relational autonomy. *Cognitive Systems Research*. DOI: http://dx.doi.org/10.1016/j.cogsys.2013.03.007.

Chalmers, D. (2006). Strong and weak emergence. In P. Clayton and P. Davies (eds.), *The Re-emergence of Emergence*. New York: Oxford University Press.

Chalmers, D. (1996). *The Conscious Mind*. New York: Oxford University Press.

Chappell, V. (1990). Locke on the ontology of matter, living things, and persons. *Philosophical Studies*, 60, 19-32.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: The MIT Press.

Chemero, A., & Silberstein, M. (2008b). Defending extended cognition. In B. Love, K. McRae, and V. Sloutsky (eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*. Austin, Texas: Cognitive Science Society.

Chemero, A., & Silberstein, M. (2008a). After the philosophy of mind. *Philosophy of Science*, 75, 1-27.

Chemero, A. (2000). Anti-representationalism and the dynamical stance. *Philosophy of Science*, 67, 625-647.

Churchland, P. S. (1993). Can neurobiology teach us anything about consciousness? *Proc. Addresses Am. Phil. Assoc*, 67, 23-40.
Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36, 181-253.
Clark, A. (2011). Finding the Mind. *Philosophical Studies*, 152, 447-461.
Clark, A. (2008). *Supersizing the Mind*. Oxford University Press.
Clark, A. (2005a). Word, niche and super-niche: How language makes minds matter more. *Theoria*, 54, 255-268.
Clark, A. (2005b). Beyond the flesh: Some lessons from a mole cricket. *Artificial Life*, 11, 233-244.
Clark, A. (2003). *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford: Oxford University Press.
Clark, A. (2001). *Mindware*. New York & Oxford: Oxford University Press.
Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58, 7-19.
Clark, A. (1998b). Time and mind. *Journal of Philosophy*, 7, 354-376.
Clark, A. (1998a). Author's reply: Review symposium on *Being There*, *Metascience*, 7, 95-103.
Clark, A. (1997). *Being* There. Cambridge, MA: The MIT Press.
Cosmelli, D., & Thompson, E. (2010). Embodiment or envatment?: Reflections on the bodily basis of consciousness. In J. Stewart, O. Gapenne, and E. A. Di Paolo (eds.), *Enaction: Towards a New Paradigm for Cognitive Science* (pp. 361-386). Cambridge, MA: The MIT Press.
Couzin, I. D., & Krause, J. (2003). Self-organization and collective behavior in vertebrates. *Advances in the Study of Behavior*, 33, 1-75.
Craver, C.F., and Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, 22, 547-563.
Craver, C.F. (2007). *Explaining the Brain: mechanisms and the mosaic unity of neuroscience*. Oxford: Clarendon Press.
Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science*, 68(1), 53-74.
Davidson, D. (1970 [1980]). Mental events. In L. Foster and J. W. Swanson (eds.), *Experience and Theory* (pp. 79-101). Amherst: University of Massachusetts Press.
De Jaegher, H., & Di Paolo, E. (2013). Enactivism is not interactionism. *Frontiers in Human Neuroscience*, 6, 1-2.

Di Paolo, E., & De Jaegher, H. (2012). The interactive brain hypothesis. *Front. Hum. Neurosci*, 6(163). DOI: 10.3389/fnhum.2012.00163.

Di Paolo, E. (2009). Extended life. *Topoi*, 28, 9-21.

Di Paolo, E., Rohde, A., & Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New Ideas Psychol*, 26, 278-294.

Donald, M. (1991). *Origins of the Modern Mind*. Cambridge, MA: Harvard University Press.

Engel, K. A. (2010). Directive minds: How dynamics shapes cognition. In J. Stewart, O. Gapenne, and E. Di Paolo (eds.), *Enaction: Towards a New Paradigm for Cognitive Science* (pp. 219-244). Cambridge, MA: The MIT Press.

Engel, K. A., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, 2, 704-716.

Fine, K. (2003). The non-identity of a thing and its matter. *Mind*, 112, 195-234.

Fodor, J. (1997). Special sciences: Still autonomous after all these years. *Philosophical Perspectives*, 11, 149-163.

Fodor, J. (1980). Methodological solipsism as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*, 3, 63-109.

Fodor, J. (1974). Special sciences (or: the disunity of science as a working hypothesis). *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science*, 28, 97-115.

French, S., & Krause, D. (2006). *Identity in Physics*: *A Historical, Philosophical and Formal Account*. Oxford: Oxford University Press.

French, S., & Ladyman, J. (2003). The dissolution of objects: Between Platonism and phenomenalism. *Synthese*, 136, 73-77.

Friston, K., Thornton, C., & Clark, A. (2012). Free-energy minimization and the dark-room problem. *Frontiers in Psychology*, 3, 1-7.

Friston, K. (2011). Embodied inference: Or, "I think therefore I am, if I am what I think". In W. Tschacher., and C. Gergomi (eds.), *The Implications of Embodiment (Cognition and Communication*), (pp. 89-125). Imprint Academic.

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

Friston, K., & Stephan, E. K. (2007). Free-energy and the brain. *Synthese*, 159, 417-458.

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, 16, 1325-1352.

Friston, K. (2002). Beyond phrenology: What can neuroimaging tell us about distributed circuitry? *Annual Review of Neuroscience*, 25, 221-250.

Froese, T., & Di Paolo, E. (2010). Modeling social interaction as perceptual crossing: An investigation into the dynamics of the interaction process. *Connect. Sci*, 22, 43-68.

Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford: Clarendon Press.

Gibbard, A. (1975). Contingent identity. *Journal of Philosophical Logic*, 4, 187-221.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton-Mifflin.

Gillett, C. (2013). Constitution, and multiple constitution, in the sciences: Using the neuron to construct a starting framework. *Minds & Machines*. DOI: 10.1007/s11023-013-9311-9.

Gillett, C. (2007a). Hyper-extending the mind? Setting Boundaries in the special sciences. *Philosophical Topics*, 35(1-2), 161-187.

Gillett, C. (2007b). Understanding the new reductionism: The metaphysics of science and compositional reduction. *The Journal of Philosophy*, 193-216.

Gillett, C. (2003). The metaphysics of realization, multiple realizability, and the special sciences. *The Journal of Philosophy*, 100(11), 591-603.

Gillett, C. (2002). The dimensions of realization: A critique of the standard view. *Analysis*, 64(4), 316-323.

Goldie, P. (2011). Grief: A narrative approach. *Ratio*, XXIV 2 June, 119-137.

Gray, W.D., and W-T. Fu. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in the world for imperfect knowledge in the head. *Cognitive Science*, 28(3), 359-382.

Griffiths, P., & Stotz, K. (2000). How the mind grows: A developmental perspective on the biology of cognition. *Synthese*, 122, 29-51.

Haken, H. (1983). *Synergetics: An Introduction. Non-equilibrium phase transition and self-organization in physics, chemistry, and biology* (third edition). Berlin: Springer.

Harris, C. B., Barnier, A. J., & Sutton, J. (forthcoming). Couples as socially distributed cognitive systems. *Memory Studies*.

Harris, C. B., Keil, P. G., Sutton, J., Barnier, A. J., & McIlwain, D. J. F. (2011). We remember, we forget: Collaborative remembering in older couples. *Discourse Process*, 48(4), 267-303.

Haugeland, J. (2002). Andy Clark on cognition and representation. In H. Clapin (ed.), *Philosophy of Mental Representation* (pp. 24-36). Oxford, UK: Oxford University Press.

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

Hawley, K. (2010). Temporal parts. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/temporal-parts/.

Hawley, K. (2006). Principles of composition and criteria of identity. *Australasian Journal of Philosophy*, 84(4), 481-493.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117, 500-544.

Hofweber, T., & Velleman, D. (2011). How to endure. *The Philosophical Quarterly*, 61(242), 37-57.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. Transactions on Computer-Human Interaction, 7(2), 174–196.

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

Hohwy, J. (2007). Functional integration and the mind. *Synthese*, 159, 315-328.

Horgan, T. (1993). From supervenience to superdupervenience: Meeting the demands of a material world. *Mind*, 102(408), 555-586.

Horgan, T. (1982). Supervenience and microphysics. *Pacific Philosophical Quarterly*, 63, 29-43.

Humphreys, P. (2008). Synchronic and diachronic emergence. *Minds & Machines*, 18, 431-442.

Humphreys, P. (1997). Emergence, not supervenience. *Philosophy of Science*, 64, 337-345.

Hurley, S. (2010). The varieties of externalism. In R. Menary (ed.), *The Extended Mind* (pp. 101-154). Cambridge, MA: The MIT Press.

Hutchins, E. (2011a). Enculturating the Supersized Mind. *Philosophical Studies*, 152, 437-446).
Hutchins, E. (2011b). Enaction, imagination, and insight. In J. Stewart., O. Gapenne., & E. Di Paolo (eds.), *Enaction: Toward a New Paradigm for Cognitive Science* (pp. 425-450). Cambridge, MA: The MIT Press.
Hutchins, E. (2008). The role of cultural practices in the emergence of modern human intelligence. *Philosophical Transaction of the Royal Society*, 363, 2011–2019.
Hutchins, E. (2005). Material anchors for conceptual blends. *Journal of Pragmatics*, 37, 1555-1577.
Hutchins, E. (1995). *Cognition in the Wild*. Cambridge: The MIT Press.
Hutto, D., & Myin, E. (2013). *Radicalizing Enactivism: Basic Minds Without Context*. Cambridge, MA: The MIT Press.
Johnston, M. (1992). Constitution is not identity. *Mind*, 101, 89-105.
Kaplan, D. M. (2012). How to demarcate the boundaries of cognition. *Biol Philos*, 27, 545-570.
Kelso, S. (1995). *Dynamic Patterns*. Cambridge, MA: The MIT Press.
Keijzer, F. (2001). *Representation and Behavior*. Cambridge, MA: The MIT Press.
Kemp, T. S. (1982). *Mammal-like Reptiles and the Origin of Mammals*. New York: Academic Press.
Kim, J. (2011). *Philosophy of Mind*, third edition. Westview Press.
Kim, J. (2006). Emergence: Core ideas and issues. *Synthese*, 151, 547-559.
Kim, J. (2005). *Physicalism, or Something Near Enough*. Princeton and Oxford: Princeton University Press.
Kim, J. (1999). Making sense of emergence. *Philosophical Studies*, 95, 3-36.
Kim, J. (1998). *Mind in a Physical World*. Cambridge, MA: The MIT Press.
Kim, J. (1993). The non-reductionist's troubles with mental causation. Reprinted (1993) in *Supervenience and Mind* (pp. 336-357). Cambridge: Cambridge University Press.
Kim, J. (1992). Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research*, 52, 1-26.
Kim, J. (1990). Supervenience as a philosophical concepts. Reprinted (1993) in *Supervenience and Mind* (pp. 131-160), Cambridge: Cambridge University Press.

Kim, J. (1989). The myth of nonreductive materialism. *Proceedings and Addresses of the American Philosophical Association*, 63(3), 31-47.
Kim, J. (1988). Supervenience for multiple domains. Reprinted (1993) in *Supervenience and Mind* (pp. 109-130), Cambridge: Cambridge University Press.
Kim, J. (1987). Strong and global supervenience revisited. Reprinted (1993) in *Supervenience and Mind* (pp. 79-91), Cambridge: Cambridge University Press.
Kim, J. (1984). Concepts of supervenience. Reprinted (1993) in *Supervenience and Mind* (pp. 53-78), Cambridge: Cambridge University Press.
Kim, J. (1973). Causes and counterfactual. *The Journal of Philosophy*, 70(17), 570-572.
Kirchhoff, M. D. (2013a). Extended cognition & the alleged causal-constitutive fallacy: In search for a diachronic and dynamical conception of constitution. *Philosophy and Phenomenological Research*, DOI: 10.1111/phpr.12039.
Kirchhoff, M. D. (2013b). Extended cognition & constitution: Re-evaluating the constitutive claim of extended cognition. *Philosophical Psychology*, DOI: 10.1080/09515089.2012.724394.
Kirchhoff, M. D. (2013c). In search of ontological emergence: Diachronic but non-supervenient. *Axiomathes*, DOI: 10.1007/s10516-013-9214-7.
Kirchhoff, M. D., & Newsome, W. (2012). Distributed cognitive agency in virtue epistemology. *Philosophical Explorations*, 15(2), 165-180.
Kirchhoff, M. D. (2012). Extended Cognition and Fixed-Properties: Steps To a 3rd-Wave of Extended Cognition. *Phenomenology and the Cognitive Sciences*, 11(2), 287-308.
Kirsh, D. (2009). Problem solving and situated cognition. In P. Robbins & M. Aydede (eds.), *The Cambridge Handbook of Situated Cognition* (pp. 264-306). Cambridge: Cambridge University Press.
Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, 73, 31-68.
Kripke, S. (1971). Identity and Necessity. In M. Munitz (ed.), *Identity and Individuation* (pp. 83-94). New York: New York University Press.
Ladyman, J., & Ross, D. (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.
Laland, K. N. (2004). Extending the Extended Phenotype. *Biology and Philosophy*, 19, 313-325.
Laland, K. N., Odling-Smee, J., & Feldman, M. W. (2000). Niche construction,

biological evolution, and cultural change. *Behavioral and Brain Sciences*, 23, 131-175.

Leonard, H. S., & Goodman, N. (1940). The calculus of individuals and its uses. *Journal of Symbolic Logic*, 5, 45-55.

Lesneiwski, S. (1916). *Powstawy ogolnej teoryi mnogosci*. *I,* Moskow: Prace Polskiego Kola Naukowego w Moskwie, Skecya matematyczno-przyrodnicza (Eng. Trans. By D. I. Barnett: *Foundations of the General Theory of Sets*. I, in Lesniewski, *Collected Works*, ed. S. J. Surma, J. Srzednicki, D. I. Barnett, and V. Rickey, Dordrecht: Kluwer, 1992, 1, 129-173.

Lewis, D. (2000). Causation as influence. Reprinted in J. Collins, N. Hall, and L. A. Paul (eds.), *Causation and Counterfactuals*. Cambridge, MA: The MIT Press.

Lewis, D. (1991). *Parts of Classes*. Oxford: Basil Blackwell.

Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

Lewis, D. (1983). Extrinsic properties. *Philosophical Studies*, 44, 197-200.

Lewis, D. (1973). Causation. *The Journal of Philosophy*, 70(17), 556-567.

Lewis, D. (1972). Psychological and theoretical identifications. *Australasian Journal of Philosophy*, 50, 249-258.

Lewis, K. (2003). Measuring transactive memory systems in the field: Scale development and validation. *Journal of Applied Psychology*, 88(4), 587-604.

Lowe, E. J. (1995). Coinciding objects: In defense of the 'Standard Account'. *Analysis*, 55, 171-178.

Lowe, E. J. (1989). *Kinds of Being*. Oxford: Basil Blackwell.

Machamer, P., Darden, L., and Craver, C.F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1), 1-25.

Markosian, N. (2004). A defense of presentism. In D. W. Zimmerman (ed.), *Oxford Studies in Metaphysics*, *Volume 1* (pp. 47-82). Oxford: Oxford University Press.

Markosian, N. (1998). Brutal composition. *Philosophical Studies*, 92, 211-249.

McDaniel, K. (2008). Against composition as identity. *Analysis*, 68(2), 128-133.

McDaniel, K. (2007). Brutal simples. In D. W. Zimmerman (ed.), *Oxford Studies in Metaphysics*, *Volume 3* (pp. 213-226). Oxford: Oxford University Press.

McKay, T. (2006). *Plural Predication*. Oxford: Clarendon Press.

McLaughlin, B., & Bennett, K. (2011). Supervenience. *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/supervenience/

McLaughlin, B. (1997). Emergence and supervenience. *Intellectica*, 25, 25-43.

McTaggart, J. E. (1908). The unreality of time. *Mind*, 17(68), 457-474.

Mellor, D. H. (1998). *Real time II*. London and New York: Routledge.

Menary, R. (2012). Cognitive practices and cognitive character. *Philosophical Explorations*, 15(2), 147-164.

Menary, R. (2010c). The holy grail of cognitivism: A response to Adams and Aizawa. *Phenomenology and the Cognitive Sciences*, 9, 605–618.

Menary, R. (2010b). Dimensions of Mind. *Phenomenology and the Cognitive Sciences*, 9(4), 561-578.

Menary, R. (2010a). Cognitive integration and the extended mind. In R. Menary (ed.), *The Extended Mind* (pp. 227-243). Cambridge, MA: The MIT Press

Menary, R. (2010). Introduction: The extended mind in focus. In R. Menary (ed.), *The Extended Mind* (pp. 1-26). Cambridge, MA: The MIT Press.

Menary, R. (2009). Intentionality, cognitive integration and the continuity thesis. *Topoi*, 28, 31-43.

Menary, R. (2007). *Cognitive Integration: Mind and Cognition Unbounded*. Basingstoke: Palgrave Macmillan.

Menary, R. (2006). Attacking the bounds of cognition. *Philosophical Psychology*, 19(3), 329-344.

Menzies, P. (2008). Counterfactual theories of causation. In *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/causation-counterfactual/

Merricks, T. (2005). Composition and vagueness. *Mind*, 144, 615-637.

Merricks, T. (1999). Composition as identity, mereological essentialism, and counter-part theory. *Australasian Journal of Philosophy*, 77, 192-195.

Mitchell, S. (2012). Emergence: logical, functional and dynamical. *Synthese*, 185, 171-186.

Moreland, R. L., & Myaskovsky, L. (2000). Exploring the performance benefits of group training: Transactive memory or improved communication? *Organizational Behavior and Human Decision Processes*, 82, 117-133.

Moreland, R. L., Argote, L., & Krishnan, E. (1996). Socially shared cognition at work: Transactive memory and group performance. In J. L. Nye and A. M. Bower (eds.), *What's Social About Social Cognition?* (pp. 57-84). Thousand Oaks, CA: Sage.

Mumford, D. (1992). On the computational architecture of the neocortex: 2. The role of cortio-cortical loops. *Biological Cybernetics*, 66, 241-251.

Nakahara, K., Hayashi, T., Konishi, S., & Miyashita, Y. (2002). Functional MRI of macaque monkeys performing a cognitive set-shifting task. *Science*, 295, 1532-1536.

Noë, A. (2006). Experience of the world in time. *Analysis*, 66(1), 26-32.

O'Connor, T., & Wong, H. Y. (2012). Emergent properties. In *Stanford Online Encyclopedia of Philosophy.* http://plato.stanford.edu/entries/ properties-emergent/

O'Connor, T., & Wong, H. Y. (2005). The metaphysics of emergence *Nous*, 39, 658-678.

O'Connor, T. (2000). Causality, mind, and free will. *Philosophical Perspectives*, 14, 105-117.

Olson, E. T. (2006). Temporal parts and timeless parthood. *Noús*, 40(4), 738-752.

Oyama, S., Griffiths, P., & Gray, R. (2001). Introduction: What is developmental systems theory? In. S. Oyama, P. Griffiths, and R. Gray (eds.), *Cycles of Contingency: Developmental Systems and Evolution* (pp. 1-11). Cambridge, MA: The MIT Press.

Pfeifer, R., Lida, F., & Bongard, J. (2005). New robotics: Design principles for intelligent systems. *Artificial Life*, 11, 99.120.

Polger, T. (2010). Mechanisms and explanatory realization relations. *Synthese*, 177, 193-212.

Polger, T., & Shapiro, L. (2008). Understanding the dimensions of realization. *Journal of Philosophy*, CV(4), 213-222.

Polger, T. (2007). Realization and the metaphysics of mind. *Australasian Journal of Philosophy*, 85(2), 233-259.

Polger, T. (2004). *Natural Minds*. Cambridge, MA: The MIT Press.

Port, R., & van Gelder, T. (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: The MIT Press.

Putnam, H. (1975). The meaning of 'meaning'. In K. Gunderson (ed.), *Language, Mind, and Knowledge*. Minneapolis: University of Minneapolis Press.

Putnam, H. (1960 [1975]). Mind and Machines. In S. Hook (ed.), *Dimensions of Mind*. New York: Collier. Reprinted in H. Putnam, *Philosophical Papers*, vol. 2, *Mind, Language, and Reality* (pp. 362-385). Cambridge: Cambridge University Press.

Rao, P. N. R., & Ballard, H. D. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79-87.

Riehle, A., Grammont, F., Diesmann, M., & Grün, S. (2000). Dynamical changes and temporal precision of synchronized spiking activity in monkey motor cortex during movement preparation. *Journal of Physiology*, 94, 569-582.

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

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

Rosen, G., & Dorr, C. (2002). Composition as fiction. In G. Richard (ed.), *The Blackwell Guide to Metaphysics* (pp. 151-174). Oxford: Blackwell.

Ross, D., & Ladyman, J. (2010). The alleged coupling-constitution fallacy and the mature sciences. In R. Menary (ed.), *The Extended Mind* (pp. 155-166). Cambridge, MA: The MIT Press.

Rowlands, M. (2010). *The New Science of Mind*. Cambridge, MA: The MIT Press.

Rowlands, M. (2009). Enactivism and the extended mind. *Topoi*, 28, 53-62.

Rowlands, M. (unpublished, ms). Perception: From extended mind to embodied phenomenology. Unpublished paper given at a symposium on Embodied Perception held at the spring 2008 meeting of the Pacific Division of the American Philosophical Association, March 18-23, 2008, Pasadena, California.

Rowlands, M. (1999). *The Body in Mind: Understanding Cognitive Processes*. Cambridge: Cambridge University Press.

Rueger, A. (2000). Robust supervenience and emergence. *Philosophy of Science*, 67(3), 466-489.

Rupert, R. (2009). *Cognitive Systems and the Extended Mind*. Oxford: Oxford University Press.

Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy*, 9¸ 625-636.

Russell, B. (1946). Reply to criticisms. In P. Shilpp (ed.), *The Philosophy of Bertrand Russell* (pp. 681-741). Chicago: Open Court.

Russell, B. (1906). Review of Hugh McColl's *Symbolic Logic and its Applications*. *Mind*, 15, 255-260.

Schewe, P. F., Stein, B., Riordon, J. (2003). A water molecule's chemical formula is really not H_2O. *Physics News Update* (The American Institute of Physics Bulletin of Physics News), 648 (July 31).

Seibt, J. (2010). Particulars. In Poli, R., and J. Seibt (eds.), *Theory and Applications of Ontology: Philosophical Perspectives* (pp. 23-55). Dordrecht, Heidelberg, London, New York: Springer.

Seibt, J. (2009). Forms of emergent interaction in general process theory. *Synthese*, 166, 479-512.

Shapiro, L. (2011). *Embodied Cognition*. London and New York: Routledge.

Shapiro, L. (2004). *The Mind Incarnate*. Cambridge, MA: The MIT Press.

Shapiro, L. (2000). Multiple realizations. *Journal of Philosophy*, 97, 635-654.

Shoemaker, S. (2007). *Physical Realization*. Oxford: Oxford University Press.

Shoemaker, S. (2001). Realization and mental causation. In C. Gillett and B. Loewer (eds.), *Physicalism and its Discontents* (pp. 74-98). Cambridge: Cambridge University Press.

Shoemaker, S. (1999). Body, self, and coincidence. *Proceedings of the Aristotelian Society* (suppl. Vol.), 73, 287-306.

Shoemaker, S. (1981). Some varieties of functionalism. *Philosophical Topics*, 12, 93-120.

Sider, T. (2007). Parthood. *The Philosophical Review*, 116, 51-91.

Sider, T. (2006). Quantifiers and temporal ontology. *Mind*, 115(457), 75-97.

Sider, T. (2001). *Four Dimensionalism: An Ontology of Persistence and Time*. Oxford University Press.

Sider, T. (1999). Global supervenience and identity across times and worlds. *Philosophy and Phenomenological Research*, 59, 913-937.

Sider, T. (1997). Four-dimensionalism. *The Philosophical Review*, 106, 197-231.

Silberstein, M. (2012). Emergence and reduction in context: Philosophy of science and/or analytical metaphysics. *Metascience*, 21, 627-642.

Silberstein, M., & McGeever, J. (1999). The search for ontological emergence. *The Philosophical Quarterly*, 49(195), 182-200.

Simons, P. (1987). *Parts: A Study in Ontology*. Oxford: Oxford University Press.

Simons, P. (1985). Coincidence of things of a kind. *Mind*, 94, 70-75.

Smart, J. J. C. (1963). *Philosophy and Scientific Realism*. London: Routledge & Kegan Paul.

Smithers, T. (1998). In time and over time. *Open Peer Commentary* on T. van Gelder (1998), “The dynamical hypothesis”. *Behavioral and Brain Sciences*, 21, 615-665.

Sober, E. (1999). The multiple realizability argument against reductionism. *Philosophy of Science*, 66, 542-564.

Spivey, M. (2007). *The Continuity of Mind*. Oxford and New York: Oxford University Press.

Sprevak, M. (2009). Extended cognition and functionalism. *The Journal of Philosophy*, 9, 503-527.

Stalnaker, R. (1996). Varieties of supervenience. *Philosophical Perspectives*, 10, 221-241.

Sterelny, K. (2010). Minds: Extended or Scaffolded? *Phenomenology and the Cognitive Sciences*, 9(4), 465-481.

Stotz, K. (2010). Human nature and cognitive-developmental niche construction. *Phenomenology and the Cognitive Sciences*, 9(4), 483-501.

Sutton, J. (2010). Exograms and interdisciplinarity: history, the extended mind, and the civilizing process. In R. Menary (ed.), *The Extended Mind* (pp. 189-225). Cambridge, MA: The MIT Press.

Sutton, J., Harris, C. B., Keil, P. G., & Barnier, A. J. (2010). The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenology and the Cognitive Sciences*, 9(4), 521-560.

Sutton, J. (2008). Material agency, skills and history: Distributed cognition and the archaeology of memory. In C. Knappett and L. Malafouris (eds.), *Material Agency* (pp. 37-55). Springer.

Thelen, E., & Smith, L. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: The MIT Press.

Theiner, G. (2013). Transactive memory systems: A mechanistic analysis of emergent group memory. *Review of Philosophy and Psychology*. DOI: 10.1007/s13164-012-0128-x.

Theiner, G. (2011). *Res Cogitans Extensa*: *A Philosophical Defense of the Extended Mind Thesis*. Frankfurt: Peter Lang.

Theiner, G., Allen, C., & Goldstone, R. L. (2010). Recognizing group cognition. Cognitive Systems Research, 11, 378–395.

Theiner, G. & O'Connor, T. (2010). The emergence of Group Cognition. In A. Corradini & T. O'Connor (Eds.), *Emergence of Science and Philosophy* (pp. 78-117). New York: Routledge.

Theiner, G. (2009). Making sense of group cognition: the curious case of transactive memory systems. *ASCS09: Proceedings of the 9th Conference of the Australasian Society for Cognitive Science*. DOI: 10.5096/ASCS200951.

Thomasson, A. (1999). *Fiction and Metaphysics*. Cambridge: Cambridge University Press.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard University Press.

Tollefsen, D. (2006). From extended mind to collective mind. *Cognitive Systems Research*, 7, 140-150.

Van Brakel, J. (2010). Chemistry and physics: No need for metaphysical glue. *Foundations of Chemistry*, 12, 123-136.

Van Cleve, J. (1990). Emergence vs. pansychism: Mind dust or magic? *Philosophical Perspectives*, 4, 215-226.

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, 21, 615-665.

Van Gelder, T. (1995). What might cognition be, if not computation? *Journal of Philosophy*, 92, 345-381).

Van Gelder, T., & Port, R. (1995). It's about time: An overview of the dynamical approach to cognition. In R. Port and T. van Gelder (eds.), *Mind as Motion: Explorations in the Dynamics of Cognition* (pp. 1-44). Cambridge, MA: The MIT Press.

Van Inwagen, P. (1994). Composition as identity. *Philosophical Perspectives*, 8, 207-220.

Van Inwagen, P. (1990). *Material Beings*. Ithaca: Cornell University Press.

Varela, F.J., Lachaux, J-P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2, 229-239.

Varela, F. J. (1999). The specious present: A neurophenomenology of time consciousness. In J. Petitot, J. F. Varela, B. Pachoud, and J-M, R. (eds.), *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science* (pp. 266-329). Stanford: Stanford University Press.

Varela, F.J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind*. Cambridge, MA: The MIT Press.

Varzi, A. (2009). Mereology. *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/mereology/

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.

Wallace, M. (2011). Composition as identity: Part 1. *Philosophy Compass*, 6(11), 804-816.

Wasserman, R. (2009). Material constitution. In *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/material-constitution/

Wasserman, R. (2004a). Framing the debate over persistence. *Metaphysica*, 5, 67-80.

Wasserman, R. (2004b). The constitution question. *Nous*, 38(4), 693-710.

Wegner, D., Erber, R., & Raymond, P. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology*, 61, 923-929.

Wegner, D. (1987). Transactive memory: A contemporary analysis of group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of Group Behavior* (pp. 185-208). New York: Springer-Verlag.

Wheeler, M. (2010). In defense of extended functionalism. In R. Menary (ed.), *The Extended Mind* (pp. 245-270). Cambridge, MA: The MIT Press.

Wheeler, M. (2005). *Reconstructing the Cognitive World*. Cambridge, MA: The MIT Press.

Wilson, R. (2010). Meaning making and the mind of the externalist. In R. Menary (ed.), *The Extended Mind* (pp. 167-188). Cambridge, MA: The MIT Press.

Wilson, R. A., & Clark, A. (2009). How to situate cognition: letting nature take its course. In P. Robbins & M. Aydede (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 55-77). Cambridge: Cambridge University Press.

Wilson, R. A. (2009). The transitivity of material constitution. *Nous*, 43(2), 363-377.

Wilson, R. A. (2007). A puzzle about material constitution & how to solve it: Enriching constitution views in metaphysics. *Philosophers' Imprint*, 7(5), 1-20.

Wilson, R. A. (2005). Collective memory, group minds, and the extended mind thesis. *Cogn Process*, 6, 227-236.

Wilson, R. A. (2004a). *Boundaries of the Mind*. New York: Cambridge University Press.

Wilson, R. A. (2004b). Realization: Metaphysics, mind, and science. *Philosophy of Science*, 71, 985-996.

Wilson, R. A. (2001). Two views of realization. *Philosophical Studies*, 104, 1-31.

Wilson, R. A. (1994). Wide computationalism. *Mind*, 103(411), 351-372.

Wimsatt, W. (2000). Emergence as non-aggregativity and the biases of reduction. *Foundations of Science*, 5, 267-297.

Wimsatt, W. (1986). Forms of aggregativity. In A. Donagan., A.N. Perovich., and M.V. Wedin (eds.), *Human Nature and Natural Knowledge* (pp. 259-291). Boston: D. Reidel.

Zimmerman, D. (1995). Theories of masses and problems of constitutions. *The Philosophical Review*, 104, 53-110.