Forthcoming in *Minds and Machines*

THE DYNAMICS OF GROUP COGNITION

S. Orestis Palermos

University *of* Edinburgh

Abstract: The aim of this paper is to demonstrate that the postulation of irreducible, distributed cognitive systems (or group minds as they are also known in the literature) is necessary for the successful explanatory practice of cognitive science and sociology. Towards this end, and with an eye specifically on the phenomenon of distributed cognition, the debate over reductionism versus emergence is examined from the perspective of Dynamical Systems Theory (DST). The motivation for this novel approach is threefold. Firstly, DST is particularly popular amongst cognitive scientists who work on modelling collective behaviors. Secondly, DST can deliver two distinct arguments in support of the claim that the presence of mutual interactions between group members necessitates the postulation of the corresponding group entity. Thirdly, DST can also provide a succinct understanding of the way group entities exert *downward causation* on their individual members. The outcome is a naturalist account of the *emergent,* and thereby irreducible, nature of distributed cognitive systems that avoids the reductionists' threat of epiphenomenalism, while being well in line with materialism.

Keywords: Distributed Cognition; Dynamical Systems Theory; Emergence; Downward Causation

> Whenever we speak about a mind, we are speaking of the processes that carry our brains from state to state…concerns about minds are really concerns with relationships between states—and this has virtually nothing to do with the natures of the states themselves.
>
> (From *The Society of Mind,* Marvin Lee Minsky)

1. INTRODUCTION

Distributed cognitive systems (or *group minds* as they are sometimes referred to in the literature) can be said to *emerge* in the following sense: They collectively exhibit socio-cognitive properties—in the form of regularities in their actual or possible behavior—that do not belong to the individual members of the group or even to their linear relations. Collective properties of this sort arise, instead, out of the ongoing reciprocal interactions between the members of the group. The goal of this paper is to argue that when this specific understanding of group cognition is viewed from the perspective of Dynamical Systems Theory (DST), it renders the postulation of the relevant group entities necessary.

Back in the 80s, Minsky (1988) and Papert (1980) suggested that individual minds may be best seen as societies of mindless, neuronal micro-agents whose complex interactions allow genuinely intelligent systems to emerge. Following this ‘society of mind’ metaphor and in order to model and understand the highly complex and distributed nature of neural networks, leading cognitive scientists have, over the years, turned to DST (e.g., Rumelhart, Smolensky, McClelland and Hinton 1986; McClelland, Rumelhart, and Hinton 1986; Varela & Singer, 1987; Turvey 1990; Thelen and Smith 1996; Port & van Gelder, 1995; Kelso 1997; Rodriguez et al., 1999; Bressler & Kelso, 2001; Thompson & Varela, 2001; Varela, Lachaux, Rodriguez, & Martinerie, 2001; Warren 2006; Spivey 2007; Dennett 1993; Varela, 1993, McClelland et al. 2010). Within philosophy of mind and cognitive science, this has come to be known as the dynamicist approach to mind and cognition.

Dynamicism has turned out to be a promising approach to individual cognitive systems. Reasoning on its basis by analogy, however, can lend plausibility to another interesting claim: Provided that the right kind of dynamic interactions are realized between individuals, distributed cognitive systems (or group minds) may emerge, just as individual minds emerge.[1]

Indeed, a growing body of literature on modeling and understanding animal collective behavior and swarm intelligence—both of which rely heavily on complex, dynamic and interactive processes between group members—attests to the utility of DST beyond the domain of individual cognitive psychology (Obuko, 1986; Niwa 1994; Parunak 1997; Li, Yang and Peng, 2009; Becco, Vandewalle, Delcourt and Poncin, 2006; Peng, Li, Yang and Liu 2010; Turnstrøm et al, 2013; Li, Peng, Kurths, Yang & Schellnhuber, 2014; Attanasi et al. 2015). As Bonabeau and Meyer (2001, 108) note, “for social insects, teamwork is largely self-organized, coordinated primarily through the interactions of individual colony members, [...] through self-organization, the behavior of the group emerges from the collective interactions of all the individuals.”

---

[1] From a dialectical perspective, the term ‘Distributed Cognitive Systems’ should be preferred over ‘Group Minds’. The reason is the common objection that minds are usually associated with consciousness, whereas groups are unlikely to enjoy consciousness over and above the consciousness of their individual members. The force of this worry, however, is not clear enough. Firstly because group consciousness may in fact be possible, and secondly because, even if impossible, its absence may not be the difference that makes the difference: Not all parts of our brains are conscious after all; accordingly not all parts of groups may need to be conscious in order to qualify as minds (for example, it may be sufficient that *some* parts, such as their individual members, are conscious). Moreover, the above objection looses considerable ground if one is willing to take the possibility of philosophical zombies seriously: If philosophical zombies are possible, then consciousness does not seem to be necessary for mindedness (see also Tollefsen 2006, fn. 11, on this point). Nevertheless, following Theiner et al. (2010, p. 379), the term ‘distributed cognitive systems’ will be here preferred over the term ‘group minds’ for the reason that no one really knows what individual minds are, which makes the idea of group minds much harder to establish. On the contrary, there is a better grasp of what specific cognitive processes (such as memory, decision-making, problem-solving, knowing, etc.) consist in, such that, should there be collective entities that manifest these cognitive processes, then we can claim that the corresponding entities may at least qualify as distributed cognitive systems.

DST has also been used in order to model and understand the emergence of human collective behavior, such as sports-team performance and rhythmic coordination (Schmidt, Bienvenu, Fitzpatrick & Amazeen, 1998; Riley, Richardson, Shockley & Ramenzoni, 2011; Duarte, Arraújo, Correia & Davids, 2012; Coey, Varlet & Richardson, 2012; Schmidt and Richardson, 2008; Duarte et al., 2013*a*; Duarte et al., 2013*b*; Dale, Fusaroli, Duran & Richardson, 2013; Richardson, Dale and March, 2014). Such studies do not always refer to human collective behavior as *cognitive* behavior. Some, however, are clearly open about employing such terminology (Marsh, Richardson and Schmidt, 2009; Cooke, Gorman, Myers & Duran, 2013). As Cooke et al. (2013, 256) note

> The term "cognition" used in the team context refers to cognitive processes or activities that occur at a team level. Like the cognitive processes of individuals, the cognitive processes of teams include learning, planning, reasoning, decision making, problem solving, remembering, designing, and assessing situations [...]. Teams are cognitive (dynamical) systems in which cognition emerges through interactions.

Admittedly, sports-team performance may not qualify as exemplary *cognitive* behavior (although the grounds for such skepticism are rarely, if ever, explicitly stated). Nevertheless, resistance to the possibility of employing the dynamicist approach for studying collective *cognitive* behavior recedes in the face of further studies within social psychology. A fast expanding body of research testifies to the promise of DST as a framework for modeling paradigmatic instances of socio-cognitive behaviors such as interpersonal communication and dialogue between adult human beings (Fusaroli, Gangopadhyay, Tylén, 2014; Fusaroli, Raczaszek-Leonardi and Tylén, 2014; Fusaroli and Tylén, 2013; Fusaroli and Tylén, 2015; Tylén, Riccardo, Bundgaard & Østergaard 2013).

Yet philosophy of mind and philosophy of cognitive science have for the most part neglected DST as a tool for motivating the existence of distributed cognitive systems. A rather surprising fact, given how receptive these disciplines are with respect to the possibility of emergent socio-cognitive collectives—*viz.,* the *hypothesis of group minds* or *distributed cognition* (Barnier et al., 2008; Heylighen et al., 2004; Hutchins, 1996; Sutton et al., 2010; Sutton, 2008; Theiner et al., 2010; Theiner, 2013a, 2013b; Theiner & O' Connor, 2010; Tollefsen & Dale, 2012; Tollefsen, 2006; Wilson, 2005).

Consider, for example, the case of Transactive Memory Systems (TMSs) (Wegner at al. 1985; Wegner 1986; Wegner et al 1991; Wegner 1995; Hollingshead 1998a, b; Hollingshead and Brandon 2003; Moreland 1999; Lewis 2003; Harris 2010). TMSs are groups of two or more individuals who interact in order to collaboratively encode, store and retrieve information. In such cases, people appear to form a group mind over and above their individual cognitive systems, because "they think about things in ways they would not alone" (Wegner et al., 1986, p. 254). As Wegner and his colleagues note, (*ibid.,* p. 256), "the observable interaction between individuals entails not only the transfer of knowledge, but the construction of a knowledge-using system that is greater than the sum of its individual member systems." Similarly, Theiner et al. (2010, p. 381) claim, "groups have the potential to display emergent cognitive properties that no individual member has, or might even be capable of having."

As the above quotes indicate, collective cognition and collective behavior in general are supposed to, in some way, *emerge*. This may initially sound unproblematic, but emergence is a notoriously slippery concept that is often associated with the idea of properties and entities that come about in mysteriously inexplicable ways. Taken out of context, for example, the above quotes from Wegner, Theiner and their colleagues are open to an interpretation according to which groups can somehow exhibit, on the basis of their members' interactions, behavior that is entirely novel compared to the (actual or possible) regularities in the behaviors of individual people, as they are described by the laws of cognitive psychology, neurobiology, chemistry and physics.[2] A claim that goes outright against the spirit of *materialism*; i.e., the very plausible claim—from a naturalistic point of view, at any rate—that all properties are, or supervene on, material properties (Stoljar 2015).[3]

This interpretation of emergence, which will be here ruled out as incorrect, is unfortunately particularly widespread, mainly due to the lack of an integrated alternative within analytic, naturalist philosophy. Emergence has been invoked within several scientific disciplines (Corradini & O'Connor, 2010) ranging over physics (e.g., Morganti, 2009; Teller, 1986), chemistry (e,g., Luisi, 2002), biology (e.g., Campbell, 1974) cognitive science (e.g., Thompson & Varela, 2001; Varela, 1993) and lately sociology as well (e.g., Sawyer, 2001, 2002, 2003). Yet so far, naturalist philosophy has not provided a systematic, mathematically inspired defense of the phenomenon that is able to resist the powerful critique of reductionism (Beckermann, Flohr & Kim, 1992; Kim, 1989, 1999).[4]

Briefly, the main objection to any approach to emergence that also accepts materialism is that it unavoidably succumbs to *epiphenomenalism*: If emergent properties and entities supervene on the physical properties of matter, then higher-level properties cannot

---

[2] 'Entirely novel' behavior here means behavior which implies some sort of substance or property dualism that would be inconsistent with what Stephan (2006, 486) calls the principle of "*Physical Monism*: Entities existing or coming into being in the universe consist solely of physical constituents. Properties, dispositions, behaviors, or structures classified as emergent are instantiated by systems consisting exclusively of physical entities." As Stephan points out, however, there are a number of other conceptions of novelty associated with emergence that are entirely compatible with Physical Monism.

[3] The term 'materialism' is often used interchangeably with the term 'physicalism', according to which all properties are, or supervene on, physical properties (for an overview, see Stoljar, 2015). The term 'physicalism' is quite ambiguous, however, and, usually, it is very closely associated with the science of physics, thereby creating the mistaken impression that all properties are reducible to the properties recognized by the language of physics. As we shall see later on, such a reading of physicalism is problematic and largely responsible for the uncharitable and mistaken interpretation of many emergentist claims. Accordingly, it is here important to insist on the subtle distinction between material and physical properties—since the latter are only a subset of the former—as well as on the distinction between the corresponding views of 'materialism' and 'physicalism'. For more details, see §4.

[4] Though note that there have been several attempts to defend emergence from a broadly 'naturalistic' perspective. See, for example, (Campbell 1974); (Humphreys 1997); (Corradini and O'Conor 2010); (O'Connor 1994); (Wilson 2013); (Wimsatt 1986; 2000) and (Stephan 1999, 2006).

have any effects over and above the causal forces of the underlying physical entities. Therefore, it is redundant to talk of 'higher-level' properties or entities, because their causal forces are in principle reducible to the causal forces of the underlying substrate. Put simply, any claim for the emergent status of a phenomenon will either have to presuppose some sort of mysterious forces that will disallow reduction, or else it will constitute a mere shortcut for referring to an aggregate of underlying entities and their causal powers that can in principle—given enough time and computational resources—be used to explain everything associated with the relevant higher-level properties and entities, without remainder.

This paper aims to demonstrate that there is a promising alternative. Specifically, there is a mathematically inspired and naturalistically respectable way to think of emergence that can resist the reductionist critique. The way to demonstrate this is to employ certain mathematical considerations from DST, in order to clarify the following set of questions: (i) *what* is the kind of interactive processes that are required for higher-level properties and entities to emerge, (ii) *why* do the relevant interactions give rise to emergence (iii) *how* should we understand emergence and (iv) *how* is it possible to avoid the dilemma of either accepting epiphenomenalism or abandoning the tenet of materialism?[5]

Extant research within cognitive science and philosophy of mind indicates that following such a dynamical approach to group cognition is not without precedent. A number of theorists, including (amongst others) Cooke et al. (2013), Warren and Fajen (2004), Coey et al. (2012), Schmidt and Richardson (2008) and Marsh et al. (2009), make regular use of DST in order to model collective behavior. Likewise, in response to question (i)—but not (ii)—Cooke et al. (2013), Duarte et al., (2013*a*; 2013*b*), Heylighen et al. (2004) and Theiner et al. (2010) have indicated that the presence of dense, mutual, or in other words, non-linear interactions between group members is a good indication that the corresponding collectives behave as unities over and above the aggregates of their individual members. Much of this literature also makes frequent use of the term 'emergence' in order to describe how such collectives come about in nature. Nevertheless, the above authors rarely seek to clarify the specific sense in which the term is used, how it relates to epiphenomenalism, or whether it is supposed to imply the irreducibility of the relevant collective entities.[6]

---

[5] In what follows, the answers to the above set of questions will be specifically concerned with the phenomenon of distributed cognition so as to provide a naturalistic approach to the emergent status of group entities and group properties (i.e., the paper's main target). *Mutatis mutandis,* however, the argument I present can be in principle applied to any case where emergence is invoked in order to understand the behavior of hierarchically organized multi-component entities.

[6] Theiner and O'Connor (2010), Theiner et al. (2010) and Theiner (2013a) provide an account of group emergence in terms of (a) the absence of intelligent design, (b) the manifestation of multiple realizability and most importantly (c) a failure of aggregativity, in Wimsatt's (1986) sense. As Wimsatt (2000) himself acknowledges, however, the problem with his approach to emergence is that it is compatible with reductionism (and thereby does not exclude the threat of epiphenomenalism). For a further critique of the above approach to emergence, see Ludwig (2015). The present account is compatible with all of the above senses of emergence, but it goes further by focusing on DST in order to provide a naturalist understanding of *downward causation* that can clearly resist the reductionist critique of epiphenomenalism.

To fill this gap, the present paper attempts to provide a *systematic* treatment of all four of the questions above. In order to answer both (i) and (ii), the focus will be on DST in order to explain the specific reasons for why non-linear interactions are considered to give rise to group entities that are irreducible to the properties of the individual members of the group (§§ 3.1 and 3.2). So far, this is a claim that has been taken for granted by both philosophers of mind and cognitive scientists, but without providing much of an explanation in its support. §3 provides two distinct arguments in support of this claim by focusing on the dynamical nature of mutually interacting systems. Moreover, in response to (iii), and again on the basis of DST, this paper provides a *naturalistic* understanding of the way in which irreducible group entities *emerge*, by detailing a dynamical account of how groups exert *downward causation* on their individual members (§§3 and 4). Finally, in response to (iv), it revisits the philosophical debate over emergence and reductionism, in order to demonstrate how the present dynamical approach to downward causation and group emergence can do away with the threat of epiphenomenalism all the while respecting the spirit of materialism (§4).[7]

2. EMERGENCE

Before looking into how DST can provide a solution to the problem of group emergence, it is helpful to be exact about what the problem that needs to be resolved is. On one hand, emergence is defined in terms of complexity: Emergent systems and properties are complex systems and regularities of behavior, respectively, that *cannot* be accounted for by (i.e., cannot be reduced to) a full and complete description of the properties of the system's components and their relations. Reductionism, on the other hand, is construed as the denial of the previous claim: Complex systems and their properties *can* be reduced to facts about the properties of their components and their relations/interactions.

Within the literature there are four possible qualifications that may be attached to any given account of emergence: 1) 'epistemological', 2) 'ontological', 3) 'weak', and 4) 'strong'.[8] Since, however, there is no general consensus (Wilson, forthcoming) on what the relation between the above kinds of emergence is supposed to be, situating the present discussion in their context would complicate matters. A better way to classify the present account of group emergence is perhaps in terms of what Stephan (1999, 2006) calls 'strong synchronic emergence'. Distinctive of this kind of emergence—as opposed to weaker forms of 'diachronic emergence' (Stephan 1999, 2006)—is the claim that the postulation of a higher-level system entails not only its *unpredictability* on the basis of the behavior of its component parts, but also

---

[7] For a number of interesting treatments of 'downward causation' see Murphy, Ellis and O'Connor (2009).
[8] For an excellent overview, see O'Connor and Wong (2012).

its *irreducibility* to them.[9] On the present account, the reason behind both of these failures is the idea that the behavior of the component parts is itself affected by the behavior of the whole on the basis of 'downward causation.'[10]

Moving beyond the existing distinctions in the literature on emergence, all naturalistically oriented approaches to the phenomenon (i.e., ones that admit that only physical matter exists) take as their starting point the supervenience thesis.

> *Supervenience*: If two events share all of their physical properties, they will share all of their social (or mental, or biological) properties as well (see, for example, Davidson, 1995; 2002).[11]

Accordingly, an entity cannot change at the social (or mental, or biological) level without also changing at the physical level. And if two events fail to share a social (or mental, or biological) property, they must also fail to share at least one physical property.

Supervenience, however, cannot be all there is to emergence, because supervenience is compatible with the claim that all higher-level properties are identical with lower-level ones, and thereby reducible to them. In technical terms, Kim (1993) has argued that if all we claim is that emergent higher level properties supervene on lower level properties, then any explanation of how, at any given instance, a higher-level property causes another higher-level property will be an explanation of how the specific underlying physical bases have an effect on each other. In other words, it will always be the case that the higher-level properties can only cause another higher- (or lower-) level property via the causal powers of the corresponding subvenient physical base. But if that's true, then, according to the "explanatory causal exclusion principle" (Kim,1989), which states that "no event can be given more than one complete and independent explanation" (*ibid.,* p. 79), higher-level properties have no genuine causal force. In other words, they are epiphenomenal (Kim, 1984; 1993, 1999).

This sounds like a serious objection to emergence. Nevertheless, not only is it not insurmountable but there might also be good reasons for resisting it. The problem with reductionism and the reason why higher-level properties are taken to merely supervene, as opposed to being identical with, the underlying lower-level properties is that higher-level

---

[9] As Stephan (1999, 49-50) points out, irreducibility entails unpredictability "since irreducible properties are *eo ipso* unpredictable in principle before their first appearance." Moreover, Stephan (52-53) notes that there can be two reasons for which a system might be irreducible: (a) Its behavior is neither micro- nor macro-scopically analyzable or (b) the behavior of its component parts does not follow from their behavior in isolation or in different constellations. The present account falls under the second version of irreducibility, which is stronger than the first, because it implies 'downward causation'.

[10] As noted above, however, it is preferable to avoid such categorization. The present account—whose distinctive feature is that it is motivated by DST—should be rather viewed as a complement to the available accounts of emergence and the choice to classify it under any of the existing categories should be left to the informed reader. It should also be noted that many of the ideas to follow appear to be in good fit with emergentist ideas expressed by Wilson (2013) and Craver and Bechtel (2007).

[11] Even though this formulation of supervenience is uncharitable to arguments for emergence, we can use it here in order to consider the reductionist argument in its strongest form. In §4, we will return to the formulation of supervenience to show how it should be amended on the face of the arguments and analysis that follows in this and the following section.

properties can be multiply realized.[12] For example, even though the higher-level property of being a couple will always supervene on two specific individuals, each given instance of that higher-level property might be realized by different kinds of individuals (e.g., it may be realized by people of the same sex, different sex, a wide range of different ages and ethnic backgrounds and so on). Still, however, as Fodor (1974) notes multiple realizability does not always entail irreducibility: If there are only a few realizing states, or if those states display some common features, reduction may still be performed unproblematically.

Reduction, however, will indeed be problematic if the underlying possible substrates of the higher-level property form a wildly unrelated list of lower-level properties. Fodor (1974, 1997) calls this kind of multiple realizability 'wild disjunction' and its distinctive feature is that, from the lower-level point of view, the underlying subvenient bases cannot be grouped together as a single (natural) kind. Think for example the higher-level property of being a family. Families always supervene on some group of specific individuals. However, each given instance of that higher-level property might be realized by entirely unrelated kinds of individuals (e.g., it may be realized by human beings, animals or even Martians). Sawyer (2001, 2002) notes that the same holds for social properties such as 'being a church', 'being an organisation', or 'being a collective movement'.[13] When this is the case, reductionism cannot explain why the higher-level properties have certain causal forces in common: How can a number of different underlying components, that appear to be entirely unrelated from the lower-level point of view, give rise to specific kinds of higher-level properties that have common causal features?

The answer that emergentists provide is that "interaction is central; higher-level properties emerge from the interactions of individuals in a complex system" (Sawyer 2001, 574). When individuals interact densely, they combine into complex dynamical system. These complex systems are regulated not only by the bottom-up causal forces governing the behavior of the individual members acting in isolation, but also by what is known as 'downward causation', arising out of the overall system as a whole.[14]

Nevertheless, the idea of downward causation remains as elusive as the general idea of emergence itself. Firstly, if downward causation is distinct from the bottom-up causal forces of the lower-level components of the system, where does it come from, and how can one even begin to account for it without parting ways with naturalism and materialism? Secondly, since reductionism is defined as the ability to account for higher-level properties on the basis of the

---

[12] For an overview on 'multiple realizability', see (Bickle, 2013).

[13] Other examples of multiply realized social (but not necessarily socio-cognitive) properties are being an 'army officer', 'being allies' and 'go to war' (for example, even ant colonies go to war: https://www.theguardian.com/environment/2016/sep/08/london-zoo-ants-1924). For more details and examples see Ruben (1985) and Tollefsen (2015).

[14] For overviews on the notion of 'downward causation' see (Campbell and Bickhard, 2011) and (Emmeche et al., 2000).

properties of the lower-level parts *and* their relations/interactions, it appears that accounting for emergence on the basis of an unqualified appeal to interactions is problematic.

Consequently, reductionists and emergentists find themselves trapped in an impasse where the former cannot explain why (wildly disjunctive) multiple realizability occurs and the latter can only suggest a vague explanation for it. To break the deadlock, it will be here attempted to throw some light on the mechanistic underpinnings of downward causation and the workings of complex systems by focusing on the conceptual framework of DST. This will lead to a naturalistic approach to emergence that can be further used to account for the irreducible nature of certain groups that form systems of distributed cognition.

3. DYNAMICAL SYSTEMS THEORY

3.1. Systems

By elaborating on van Gelder (1995) and Beer (1995), it is possible to invoke DST (and especially the concept of a *coupled system*) in order to provide a plausible—because somewhat more stringent than the rest—mathematically informed version of the hypothesis of extended cognition (Palermos 2014). The same approach and arguments involved thereof can be employed in order to argue for the emergence of distributed cognitive systems. This is not surprising, given that the only difference between the extended and distributed cognition hypotheses is that, in the latter case, the system extends to include not only artifacts but other cognitive agents as well. Before looking at the arguments themselves, it will be helpful to go through some of the technical terms that will figure in the discussion as well as restate the reasons for which DST is pertinent to the present debate.

Dynamical modeling—the part of mathematics that is concerned with understanding natural phenomena by providing abstract dynamical models for them—and *Dynamical Systems Theory*—the branch of theoretical mathematics that is concerned with the properties of such abstract dynamical models—are the backbone of the most successful theories within physics and chemistry. Moreover, the conceptual and modeling resources of DST are having a growing impact within cognitive science both with respect to modeling *individual* cognitive processes (e.g., Beer, 1995; Bressler & Kelso, 2001; Chemero, 2009; Dale & Spivey, 2006; Froese, Gershenson, & Rosenblueth, 2013; van Gelder, 1995; Kelso & Engstrøm, 2008; Palermos 2014; Spivey & Dale, 2006; Thompson & Varela, 2001; Spivey, 2007; Varela et al., 2001; Port and van Gelder, 1995) as well as *collective* ones (e.g., Cooke et al. (2013), Warren and Fajen (2004), Coey et al. (2012), Schmidt and Richardson (2008), Marsh et al. (2009), Dale, Fusaroli, Duran and Richardson (2013)).

In addition to its prominence within the scientific enterprise, DST is also pertinent to the present debate, because dense interactions and the complex systems they give rise to are central to understanding the emergence of distributed cognitive systems (Cooke et al. (2013), Duarte et al., (2013*a*), Duarte et al. (2013*b*), Heylighen et al. (2004), Theiner et al. (2010), Fusaroli, R., Rączaszek-Leonardi, & Tylén, (2014), Fusaroli, Gangopadhyay, & Tylén (2014)), and DST is the best—and possibly the only—framework for studying such interactively rich behavior.[15]

Two distinct reasons motivate, therefore, the focus on DST as a tool for understanding the emergence of group cognition: (1) Continuing its impressive potential to provide successful explanatory models within the natural sciences, DST is currently demonstrating a growing momentum within the general field of cognitive science too. (2) Several cognitive scientists have indicated that dense mutual interactions are crucial for the emergence of group cognition, and DST is an ideal candidate for generating explanatory models for them.

Now, the best way to introduce the technical terms that will figure in the arguments to follow is to start with the fundamental concept of a 'system'.[16] Systems are sets of interdependent elements, objects, entities, or items standing in interrelations on the basis of specific processes they take part in and give rise to, thereby forming a unified whole. It is important to note that an element, object, entity, or item can be part of several systems at the same time, depending on the kind of processes it engages in. Thus, whether some object counts as a component of a system always depends on the phenomenon under study and, more in particular, on the processes that are thought to give rise to the relevant phenomenon.[17]

Dynamical Systems Theory Terms	**Definition**
State variables, *x*	The values of the changing aspects of the system. Also, the system's output: the way the system behaves in response to changes in its parameters.
Dynamical Law, *L*	A set of (usually) differential equations that regulates the change of the state variables
Trajectory	A sequence of states generated by the dynamical law, starting from some initial state x_0
State space	The set of all possible values of the system's state variables
Flow	The set of all possible trajectories through every point in the system's state space
Limit sets	Sets of points towards which the system will always converge over time

---

[15] For a general introduction to Dynamical Systems Theory see (Abraham, Abraham, & Shaw, 1990).

[16] For ease of reference, table 1 includes the definitions of most of the terms that figure in the discussion to follow. They are listed in the same order they appear in the main text (starting with the most basic terms and moving on to the more complex ones).

[17] Craver and Bechtel (2007) make similar remarks in the context of a discussion on mechanisms and downward causation.

Attractor	A limit set that gravitates trajectories passing through all nearby states
Basin of attraction	A set of initial states that converge to a given attractor
Transient	The portion of a trajectory that is found within a basin of attraction but which does not lie in the attractor itself
Repellor	An unstable limit set whose nearby trajectories diverge from it
Non-autonomous system	A system whose dynamical law depends not just on the values of its state variables, x, and fixed parameters, u, but also on the values of some set of changing parameters u(t)
Fixed parameters, *u*	The values of the dynamical law that are determined by the internal features of the system (e.g. the material it is made of), or the background conditions it operates in
Changing parameters, *u(t)*	The input to a non-autonomous system
Phase portrait	The graphical representation of attractors, repellors and basins of attraction (i.e., the different phases) the system can enter into
Bifurcation	A qualitative change (e.g., the appearance or disappearance of attractors and/or repellors) in the system's phase portrait, caused by a change in its parameters (either fixed or changing)
Coupled system	A system consisting of two mutually interacting non-autonomous systems, whose parameters function as some of the state variables of the other, and *vice versa*
Parameter space	The set of all possible combinations of values for all the different parameters of a given system

Table 1: Definitions of relevant dynamical systems theory terms

Moving on to the main features of systems, every dynamical system is characterized by a set of *state variables, x,* and a *dynamical law, L*—a set of (usually) differential equations—that regulates the change of those *state variables* across time. Starting from some initial state x_0 the law, *L*, generates a sequence of states, which is called the *trajectory* of the system. The set of all *trajectories* through every point in the *state space* is called the *flow.* In general, DST is primarily concerned with the geometrical structure of the entire *flow* of the system: i.e., the geometrical or topological properties of all the possible behaviors the system may exhibit across time.

The most important behavior of any system is the convergence to *limit sets*: i.e., sets of points that are unaffected by the *dynamical law* in that regardless of what the initial state x_0 of the system is, the system will always end up to one of them as time goes to infinity. *Limit sets* are important, because they can be used to understand *the long-term regular behavior—i.e., the properties—of the system*.

Some *limit sets,* called *attractors,* gravitate *trajectories* passing through all nearby states. The set of initial states that converge to a given *attractor* is called its *basin of attraction* and the portions of the *trajectories* that are found within a *basin of attraction*, but which do not lie in the attractor itself, are termed *transients*. We can use these concepts in order to understand why *attractors* are doubly important*:* Firstly, *attractors* govern the long-term behavior of the system as the system tends to converge to them once *transients* have passed, regardless of what the system's initial state is. Secondly, once the system has entered an *attractor*, it will tend to remain there—even if disturbed—due to the attractor's basin of attraction.

Conversely*, repellors* are *limit sets* that are unstable in that some nearby *trajectories* diverge from them. In result, they cannot retain—as it were—the system in their state, even if the system is only slightly disturbed.

The *state space* of any given dynamical system contains multiple *repellors* and *attractors*. Let us repeat that these *limit sets* are responsible for the regularities in the system's behavior and can be used in order to capture the system's properties.

Now, most of the systems that exist in nature are *non-autonomous* in that their behavior is governed not just by the changing values of their *state variables, x(t),* and some *fixed parameters*, *u* (whose values remain constant across time) but also by the values of some *changing parameters, u(t).* The values of a system's parameters and how they can affect the system is very important for distinguishing between stable and unstable systems. If changing the values of the parameters produces small changes in the resulting *flow,* we can say that the system is structurally stable.[18] Other systems, however, are unstable in that very small changes in the values of their *parameters* can produce substantial changes in their *flow.* The result is *phase portraits* that are qualitatively different from the initial one. For example, new *attractors* may appear and old *repellors* may disappear. Such *qualitative changes* in the system's flow are called *bifurcations* and they are very important, because they signify the emergence of new systemic properties.

Further on *parameters*, it is also important to distinguish between the above two kinds (i.e., *changing* and *fixed parameters*) with respect to their representational roles. On one hand, *constant parameters, u,* refer either to one of the internal features of the system that may be manipulated (but which remain fixed during the system's operation) or to the stable background conditions the system operates in. On the other hand, *changing parameters, u(t),* represent the *inputs* to the system. These inputs might originate from the system's dynamical environment or some other well-defined system that causally affects the system under study.

While keeping the above in mind, it is interesting to see what happens when these *inputs* (i.e., *changing parameters u(t)*) do not just originate from the system's dynamical environment, but from another system that the system under study is non-linearly related with on the basis of *mutual interactions* that arise out of feedback loops between the two. These mutual interactions bring the two systems together into an overall *coupled system,* which means that some of the *changing parameters u(t)* of each system function as state variables of the other

---

[18] "Limit sets and basins of attraction may deform and move around a bit, but the new flow will be qualitatively similar (i.e., topologically equivalent, or *homeomorphic*) to the old one" (Beer, 1995, p. 180).

and *vice versa.* Typical examples of such systems include two mutually interconnected pendulums, the Watt governor and a rotation engine, arguably certain cases of cognitive agents employing their artifacts and, possibly, groups of interacting individuals.

Imagine two non-autonomous cognitive systems, Rudy and Lulu, who form parts of a transactive memory system:

> Suppose we are spending an evening with Rudy and Lulu, a couple married for several years. Lulu is in another room for the moment, and we happen to ask Rudy where they got that wonderful staffed Canadian goose on the mantle. He says "we were in British Columbia…," and then bellows, "Lulu! What was the name of that place where we got the goose?" Lulu returns to the room to say that it was near Kelowna or Penticton—somewhere along lake Okanogan. Rudy says, "Yes, in that area with all the fruit stands." Lulu finally makes the identification: Peachland (Wegner et al., 1985, 257).

Commenting on the dynamics of such a case, Beer (1995, 182) notes that we "cannot overemphasize the fundamental role that feedback plays in this relationship." Any act of communication that Lulu takes affects Rudy, which, in turn, affects back herself via the feedback she continuously receives from him, and *vice versa.* Each of the two dynamical systems is continuously shaping the *flow* of the other (possibly not just in a quantitative way, but also qualitatively if the coupled *parameters* of the receiving system exceed *bifurcation* points in the receiving system's *parameter space*), thereby drastically influencing its subsequent *trajectory*. During their discussion, the various ideas, expressions and behavior that Rudy and Lulu exchange allow them to not only elicit richer individual memories in a faster way (a mere quantitative change), but also enjoy memories that they wouldn't normally have access to (a *bifurcation*). Via this process of interactive cueing they can move sequentially toward the retrieval of memory traces that, regardless of whether their existence is known to both of them or not, would normally be unavailable to either of them, were they to act as isolated individuals.

Accordingly, given (1) that the cognitive process of recollecting arises out of this kind of direct, reciprocal dependence between Rudy and Lulu and (2) the definition of systems as sets of interdependent elements standing in interrelations on the basis of specific processes they participate in and regular behaviors they give rise to, we can view the two coupled non-autonomous systems, Rudy and Lulu, as a single unified non-autonomous cognitive system, Rulu. This is because some of the cognitive properties we are interested in, such as the regular recollection of autobiographical beliefs on the basis of transactive memory processes, belong to the overall system as a whole.

Moreover, the *state variables* of this overall unified system are the union of the free *state variables* of the two subsystems (i.e., the *state variables* that do not participate in the coupling) along with a set of *collective variables* (alternatively, *order parameters*). '*Collective variables*' is a central concept of self-organization and of primary importance in the constitution of the overall system, because they govern the interactions of the overall system as a whole. The way they do it is by constraining or "prescribing" the behavior of the component parts… "'enslaving' them, as it were, so that they no longer have the same behavioral alternative open to them as would be the case if they were not interdependently linked in the system"(Thompson & Varela, 2001, p. 421).

Order parameters/control variables, in other words, are responsible for the global characteristics of the overall system that govern or constrain local interactions. They give rise to global patterns of behavior that determine the action of the individual members in the sense of 'allowing' to the system's local parts only *some* of the possible alternative behaviors they could exhibit were they not coupled (and thereby not parts of the overall system).

Collective variables are therefore responsible for the coordinated regular behavior of all or at least several components at the same time. Importantly, such coordinated regular behavior can only be described and conceptualized as properties (i.e., *attractors, repellors,* etc.), in the phase portrait of the overall system as a whole.

McGrath et al. (2000, p. 100), for example, suggest modeling conflict in a small group as a *collective variable* that depends on the *changing* contextual *parameter* of external stress:

> [O]ne group (the alpha team) may have a single, stable, fixed point attractor of moderate conflict. This is the value for conflict the group settles into and maintains. This configuration holds over a range of values for external stress (a contextual parameter) on the group. At very high levels of stress, however, a new pattern may emerge, with two unstable attractors of either very high conflict or very low conflict. Another group (the beta team) may have a stable periodic attractor for conflict—a consistent pattern of increasing, then decreasing, conflict—that persists over a wide range of stress levels. At very high stress, however, the beta team shifts to a single, stable, fixed point of high conflict.

Likewise, a similar model may be provided for a TMS that is presented with a relevant question. Under conditions of sobriety (represented as one of the *fixed parameters, u,* of the individual *systems*) the group engages in conversation—a *collective variable*—until it reaches a successful answer. Once this happens and if the dyad is not presented with further relevant *input*, its transactive communication stops, signifying it has converged to one of its *attractors*, wherein the value of the *collective variable* is at minimum. A change, however, in the individual *fixed parameters*, *u,* when say both members are under the influence of alcohol, can make the *collective variable*—i.e., the transactive communication—enter a never-ending state of *transients*. When the TMS is in that state, its interactive cueing never converges to an *attractor* that represents the successful retrieval of any given answer.[19]

Such modeling suggestions are motivated by the increasing trend within cognitive

---

[19] In (Arrow et al., 2000), the authors go through several examples of how DST could be used to model the behavior and properties of groups in terms of collective variables. Some of the suggested examples include the quantity or rate of production of the group's product; the quality of the group product; the temporal features of conflict, such as speed of escalation and de-escalation; the discrepancies between member behavior and shared normative expectations; the development of group task strategies; leadership structures; patterns of communication, and so on. For more examples and details, see pp. 134-137 and pp. 148-156.

science to employ DST as a tool for modeling and understanding collective behavior. Nevertheless, specific aspects of the above systems may raise worries with respect to their amenability to a DST treatment. For example, one possible objection is that groups of people, including TMSs, communicate via language, which is discretely symbolic (i.e., non-continuous), such that DST is not the best tool for modeling their behavior.

It should be kept in mind, however, that face-to-face communication is not only verbal but also non-verbal and paraverbal. Think, for example, of the multitude of meaning conveyed by face expressions, body language and the tone, pitch and pacing of voices as well as nonlinguistic sounds of agreement or disagreement that hearers make during their interlocutor's speech acts. Such dynamic aspects of interpersonal communication are already being successfully modeled on the basis of DST (Raczaszek-Leonardi & Kelso, 2008; Fusaroli, Raczaszek-Leonardi and Tylén, 2014; Fusaroli, Gangopadhyay, Tylén, 2014; Fusaroli and Tylén, 2013; Fusaroli and Tylén, 2015; Tylén, Riccardo, Bundgaard & Østergaard 2013), rendering the DST framework particularly promising for modeling TMSs, or any other group, wherein collective behavior arises on the basis of members' communication.[20]

3.2. Two arguments from Dynamical Systems Theory

Before moving further, it is helpful to repeat the upshot of the previous subsection: DST indicates that the coupling of non-autonomous dynamical systems into one unified system can give rise to properties (i.e., regularities in actual or possible behavior) that go beyond the sum of the properties the individual systems can produce on their own. In other words, some of the geometrical properties (such as the *attractors* and *repellors*) of the resulting *flow* will not be attributable to any subsystem alone but only to the overall system as a whole. Moreover, when

---

[20] An anonymous referee points out that, instead of outlining how cognitive science can employ the main concepts and techniques of DST in order to model the behaviour of distributed cognitive systems such as TMSs, it would be preferable to actually offer such a detailed model. While offering such a model would no doubt add to the plausibility of the paper's overall argument, exploring and developing such a model is beyond both the scope of the present paper and the available space. The present paper aims to demonstrate that should DST be a promising tool for modelling distributed cognitive systems such as TMSs, then it is possible to provide a naturalistically respectable argument for the emergent, irreducible nature of distributed cognitive systems as well as a rigorous understanding of the downward causation that such collective systems exert on their individual members (for more details see sections 3.2, 3.3 and 4). Similarly, offering a successful DST model of a distributed cognitive system such as a TMS would add to the prospects of DST as a successful tool for modelling collective systems in general and distributed cognitive systems in particular. Nevertheless, as the above notes and many of the studies that are cited in the introduction of the paper indicate, there is a fast growing body of research (e.g., Raczaszek-Leonardi & Kelso, 2008; Fusaroli, Raczaszek-Leonardi and Tylén, 2014; Fusaroli, Gangopadhyay, Tylén, 2014; Fusaroli and Tylén, 2013; Fusaroli and Tylén, 2015; Tylén, Riccardo, Bundgaard & Østergaard 2013) that has already started employing DST concepts in order to model collective cognition and behaviour and, indeed, several of these studies (e.g., Schmidt, Bienvenu, Fitzpatrick & Amazeen, 1998; Coey, Varlet & Richardson, 2012; Schmidt and Richardson, 2008; Duarte et al., 2013*a*; Duarte et al., 2013*b*; Richardson, Dale and March, 2014) have been successful in providing considerably detailed DST models of collective phenomena such as sports team performance and rhythmic coordination. Within the literature, therefore, there is growing evidence attesting to the promise of DST as a successful tool for modeling the behavior of distributed systems such as TMSs, which, in addition to strengthening the present paper's overall argument, offers strong incentive for carrying out future empirical studies in this exciting direction.

these properties refer to regularities of behavior that one would readily classify as *cognitive* (such as the recollection of autobiographical memories), then the underlying components—whose mutual interaction generates the relevant (cognitive) behavior—may be said to form proper parts of an overall (distributed) *cognitive* system that consists of all of them at the same time.

It is important to note that this is not to say that the overall system exhibits forms of behavior that somehow go against, or could not be conceived of by merely considering, the properties of the component parts (when prompted, for example, individuals do, occasionally, enjoy old autobiographical memories). Rather, what it means is that the overall system restrains the behaviors of the component parts to just a few of all the possible behaviors they would tend to exhibit were they to act on their own (for example instead of replying "I don't know" or going to look for the answer in an old photo album, the TMS tends to regularly engage its members into transactive communication processes). In effect, this restrained behavior gives rise to *new regularities of behavior* (such as the regular successful recollection of old autobiographical memories) that the component parts do not bear on their own. Instead, such regularities of behavior originate from and thereby belong to the overall system as a whole.

In other words, the mutual interactions between the individual members of the group give rise to new systemic properties (such as periodic patterns of conflict, or the efficient recollection of shared but, normally, individually inaccessible memories) that do not belong to any of the subsystems alone, but to the overall group, taken as a coupled system. Moreover, as noted in the beginning of the previous section, system individuation does not depend on any physical boundaries, but, instead, on the processes and relevant properties one is interested in, and which emerge out of component interactions. Taken together, these two points provide us with a first reason for thinking that the postulation of coupled systems and thereby distributed cognitive systems (when the distributed process we are focusing on is a cognitive one) is far from redundant.

Put another way, in cases where individuals generate cognitive processes by mutually interacting with each other, the postulation of a single distributed cognitive system brings explanatory value: Specifically the postulation of the distributed cognitive system is necessary with respect to the explanation of certain systemic properties, which we would otherwise be at a loss how to account for. Accordingly, distributed cognitive systems are not open to the common eliminativist line that Xs do not exist because our best explanations are not committed to the existence of Xs (i.e., that positing Xs does no explanatory work).[21] Postulating distributed cognitive systems is therefore necessary. We can call this the 'systemic properties' argument for the postulation of distributed cognitive systems (see also Palermos 2014).

---

[21] Rupert has pressed this objection against group cognition in a number of places (2005, *forthcominga*, *forthcomingb*).

This is a self-standing argument for the postulation of distributed cognitive systems. DST, however, can also provide an additional argument to the same effect. To see how this second argument goes, recall first that a non-autonomous system's *changing parameters, u(t),* refer to the system's *inputs,* which may originate from the dynamical environment, or some other well defined system. Accordingly, when we have two causally—but not mutually—dependent systems, the input refers to the effects of the affecting system on the affected system and, conversely, the *output* refers to the affected system's reaction (i.e., the system's behavior) to its *input*. Moreover, assuming that there is only one-way causal dependence, the affected system's output has no direct bearing on the affecting system's dynamics, and will thus be represented only as quantitative changes in one, or more, of the affected system's *state variables*. In cases of two causally (but not mutually) dependent systems, then, we can clearly tell the behaviors of each system apart in terms of distinct *inputs* and *outputs* from the one system to the other.

In cases of non-autonomous coupled systems, however, where some of the *changing parameters u(t)* of each system function as state variables of the other and *vice versa*, talk of inputs and outputs is inapplicable. The reason, as we saw Beer (1995, p. 182) pointing out, has to do with the fundamental role that feedback loops play in this relationship. For instance, the way each individual in the group is going to be affected by the other is *partly determined* by the individual member itself: The effects it receives are not exogenous to itself and therefore cannot be properly conceptualized either as its *input* or as the affecting component's *output*. Put another way, in cases where two non-autonomous systems mutually interact on the basis of feedback loops, there will be a *causal amalgam* between the contributing units that resists their decomposition into two separate systems on the basis of distinct *inputs* and *outputs*. Consequently, since we cannot disentangle the behavior of the contributing members in this way, we must accept they constitute an overall system comprising of both of them. We can call this the 'ongoing feedback loops' argument for the postulation of coupled systems, in general (see also Palermos 2014). Additionally, when the relevant feedback loops generate behavior that would be readily classified as cognitive, the 'ongoing feedback loops' argument can also act as a further argument for the postulation of distributed *cognitive* systems in particular.

Therefore, it is possible to provide two distinct arguments for the postulation of distributed cognitive systems. First, the cognitive properties that arise out of the reciprocal, non-linear interactions of two or more individuals cannot be attributed to any of the contributing members or their sum, but to their coupled system as a whole. Accordingly, we *have to* postulate the overall distributed cognitive system. (Alternatively, distributed cognitive systems are necessary for accounting for such systemic properties, so they cannot be ontologically eliminated). Second, in cases where individuals generate cognitive processes on the basis of ongoing feedback loops between them, there is a dense, non-linear causal interdependence that cannot be decomposed in terms of distinct inputs and outputs from the one agent to the other (the reason being that the effects of each individual to the other are not entirely endogenous to themselves, and vice versa). Accordingly, we *cannot but* postulate the overall distributed cognitive system that those individuals form part of.

The above two arguments also reveal a broader point. They both indicate and explain why the presence of *ongoing mutual interactions* (and the resulting non-linear relations) between the members of a group can be treated as the *criterion* by which we can judge whether the relevant group can qualify as *a coupled system* in its own right. In practice, this means we cannot claim that asking for directions from a stranger on the street or receiving testimony in the court of law can give rise to a distributed cognitive system. In such cases *there are no non-linear relations* between the cognitive processes of the involved individuals. The cognitive processes of the individual that produces the relevant information are not mutually interrelated with the cognitive processes of the individual that receives the information. Instead, there is only one-way, linear dependence between the individuals under consideration; the way the speaker formulates the information she delivers is entirely independent of the recipient's cognitive processes. On the contrary, in the case of TMSs (Wegner et al., 1985; Wegner, 1986; Sutton 2008; Sutton et al. 2010), the completion of the relevant cognitive task involves dense feedback loops between the participating individual members, suggesting that the criterion of ongoing mutual interactions is satisfied.[22] Accordingly, in such cases, we can indeed talk of the presence of an overall distributed cognitive system that consists of all the participating individuals at the same time.

Cognitive systems are therefore genuinely distributed only in those cases where we face a task (i.e., a process—and remember we individuate systems on the basis of the processes we are interested in) that we would intuitively like to call a cognitive one, and which is accomplished on the basis of ongoing mutual interactions between two or more individuals. Otherwise—if the assumed cognitive task is not completed on the basis of ongoing mutual interactions, but merely on the basis of one-way dependences between the participating individuals—we may only talk of cognition as being socially embedded (but not distributed).

3.3. An overarching worry: Identifying cognitive systems and processes.

It is important to address an overarching worry concerning the cognitive nature of the distributed systems and processes advocated above. In doing so, it is helpful to formally restate the previous section's overall argument:

---

[22] Prominent ethnographers and philosophers of science would also provide the examples of several scientific research teams (Knorr-Cetina, 1999, Nersessian, 2006; Giere, 2002*a,* 2002*b*)

Argument for Distributed Cognition (ADC)

> P1: A process Δ is brought about on the basis of mutual interactions between the members of a group.
> P2: According to the 'systemic properties' and 'ongoing feedback loops' arguments, when component parts mutually interact with each other in order to bring about some process Π, there exists (with respect to Π) an overall system that consists of all of them at the same time.
> C1 (From P1 and P2): With respect to Δ, the underlying group constitutes a distributed system that consists of all the interacting individuals.[23]
> P3: Δ is a process that, on the basis of common sense intuitions, we would readily classify as *cognitive*.
> C2 (From C1 and P3): With respect to Δ, the underlying group constitutes a distributed *cognitive* system that consists of all the interacting individuals.

A possible worry with this line of reasoning is that it may commit what Adams and Aizawa (2008) have dubbed the systems version of the 'coupling-constitution' fallacy. This fallacy is usually associated with arguments for the hypothesis of extended cognition, but it is possible to raise similar concerns with respect to arguments for distributed cognition. Following Adams and Aizawa's original formulation (Adams & Aizawa, 2008, p. 92), the fallacious move is supposed to unfold in two steps: The first is to move from the observation of some sort of causal connection to the claim that several individuals form a system (this is the conclusion C1 drawn from P1 and P2, above). The second step is a tacit shift from the hypothesis that something constitutes a system to the hypothesis that it is an instance of a distributed *cognitive* system. As Adams and Aizawa suggest, this second step is fallacious: "It simply does not follow from the fact that one has identified an X system in terms of a causal process of type X that that process pervades every component of the system" (ibid., p. 125).

Granting to Adams and Aizawa that the above move is indeed fallacious, it should be noted that ADC does not make this move. Committing the fallacy that Adams and Aizawa have in mind in the case of distributed cognition would take the following form: Individual members, A, B, C, D... of a group, G, interact mutually with each other on the basis of their individual-level *cognitive* processes. Therefore, since the individual members of the group manifest cognitive processes, G should also count as a distributed *cognitive* system. Put another

---

[23] In other words, the underlying group includes as its proper parts the cognitive systems of all the interacting individuals. An anonymous referee notes that this raises the question of how the different, interacting levels of cognitive systems stand in relation to each other. Briefly, the behaviour of the distributed cognitive system supervenes on the behaviour of the underlying individuals' cognitive systems. At the same time, the behaviour of the individuals' cognitive systems is affected, via downward causation, by the activity of the distributed cognitive system they are parts of. For more details, see section 4.

way, the fallacious move would be to claim that the overall system is *cognitive*, because its members mutually interact with each other on the basis of *individual-level cognitive* processes. But this is not how ADC works.

While ADC appeals to the mutual interactions of the individual members of the group (P1) in order to claim, on the basis of the 'systemic properties' and 'ongoing feedback loops' arguments (P2), that there is an overall distributed system G (C1), it does not hold G to be cognitive, because of the cognitive processes manifested by the individual members of the group. This is an independent claim that is established on the basis of P3—i.e., that, intuitively, the overall processes that arises on the basis of the individual members of the group is one that would be readily classified as cognitive. Independently of whether this is a valid move, it keeps ADC free of the fallacious move that Adams and Aizawa worry about.

Let us then focus on P3 instead. How safe is it to claim that a distributed system is *cognitive* because the mutual interactions of the members of the group give rise to an overall process that could be readily classified as cognitive? There are two possible objections to this move.

First, claiming that a group of people may qualify as a distributed *cognitive* system because there exists a *cognitive* process whose realization depends on the mutual interactions of the members of the group sounds similar to Theiner's (Theiner et al. 2010, Theiner 2013a) *Social Parity Principle*:

> If, in confronting some task, a group collectively functions in a process which, were it done in the head, would be accepted as a cognitive process, then that group is performing that cognitive process.

Accordingly, it is fair to wonder whether ADC has anything to add to this. Moreover, if ADC has nothing to add to the Social Parity Principle, but simply constitutes a detailed reiteration of it, it is important to ask whether the Social Parity Principle itself can stand as an adequate argument in support of distributed cognition.

In response, the Social Parity Principle is neither identical to ADC nor sufficient as an argument for distributed cognition. The two arguments are similar in that they both appeal to common-sense intuitions in order to establish which processes may qualify as cognitive. Granting for the time being that this appeal to common-sense intuitions is unproblematic (this is the second worry one may have and we shall return to it soon), it should be noted that, beyond this appeal to common-sense intuitions, ADC and the Social Parity Principle come apart. If the Social Parity Principle were to be taken as a self-standing argument in its own right (note that Theiner et al. (2010, 384) suggest otherwise) it is not sufficient to motivate distributed cognition. This is because, even though the Social Parity Principle may help judge whether a given process can qualify as a *cognitive* process, it cannot further establish its collective nature. It provides no argument for the additional claim that the relevant cognitive process is irreducible to the sum of the cognitive processes possessed by the individual members of the group. Indicatively, note how the antecedent of the Social Parity Principle already presupposes that the relevant process is *collectively* performed. This is problematic because, contrary to whether a process may qualify as cognitive, arguing for its irreducibly collective nature cannot be decided on the basis of common-sense intuitions but must instead be explicitly defended on a mechanistic basis. This is precisely the argumentative role performed by the 'systemic properties' and 'ongoing feedback loops' arguments, invoked in P2 of ADC.[24] ADC therefore adds to the Social Parity Principle by offering an independent, mechanistic rationale for why some process (independently of whether it is cognitive or not) is collectively performed by some distributed system.

Yet another worry remains, which can be directed against both the Social Parity Principle and ADC. The worry concerns the appeal to common-sense intuitions in order to argue that a given process may qualify as cognitive. In order to establish that the relevant collective process is a collective *cognitive* process, both the Social Parity Principle and ADC require that we be willing to accept—on the basis of common sense intuitions—that the relevant process could be readily classified as cognitive (i.e., P3 of ADC).[25]

Some authors suggest that this move is worrying (Huebner (2013), Ludwig (2015)) because it appears to be suspiciously close to behaviorism. The reason is that ADC (via P3) and the Social Parity Principle appear to qualify a process as cognitive merely on the basis of *behavior* that could be classed as cognitive. It should become immediately obvious, however, that this is an uncharitable characterization of the above arguments. For while they do take as their starting point that some process may be classed as cognitive, because it exhibits behavior that we would normally classify as such, contrary to behaviorism, they do not assume that all there is to cognition is behavior alone.[26] Interestingly, Graham (2015) notes that this difference between the philosophical current of behaviorism and employing intelligent behavior as evidence for the presence of cognition was in fact clearly recognised by Sellars, a long time ago:

> Wilfred Sellars (1912–89), the distinguished philosopher, noted that a person may qualify as a behaviorist, loosely or attitudinally speaking, if they insist on confirming "hypotheses about psychological events in terms of behavioral criteria" (1963, p. 22). A behaviorist, so understood, is someone who demands behavioral evidence for any psychological hypothesis. For such a person, there is no knowable difference between two states of mind (beliefs, desires, etc.) unless there is a demonstrable difference in the behavior associated with each state. Consider the current belief that it is raining. If there is no difference in my behavior between

---

[24] Theiner (forthcoming) distinguishes between several approaches to group cognition. ADC would fall under GC6, i.e., "the Dynamical Stance."

[25] Though note a significant difference: the Social Parity Principle holds that the relevant process is cognitive, because it would count as cognitive were it to be performed *within the agent's head*. P3 of ADC does not put forward such an additional criterion regarding the locus of individual cognition. This is an advantage of ADC, because as Ludwig (2015) argues, this additional appeal to brain-bound cognition invites a number of problems.

[26] For an overview of behaviorism, see Graham (2015).

> believing that it is raining and currently thinking that the sun is bright, there is no grounds for attributing the one belief to me rather than the other. The attribution is empirically unconstrained. Arguably, there is nothing truly exciting about behaviorism loosely understood. It enthrones behavioral evidence, an arguably inescapable premise in not just psychological science but in ordinary discourse about mind and behavior. Just how behavioral evidence should be 'enthroned' (especially in science) may be debated. But enthronement itself is not in question. Not so behaviorism the doctrine.

In fact, the distinction between 'attitudinal' and philosophical behaviourism is particularly important, for, in its absence, it would render most of contemporary cognitive science behaviourist. In the absence of a 'mark of the cognitive', cognitive scientists have no other way to distinguish between the presence and absence of cognition, other than by employing, in essence, what Sellars calls 'attitudinal' behaviourism.[27] Wilson (2001) puts the same point in the following way:

> In order for something to have a mind, that thing must instantiate at least some psychological processes or abilities. Rather than attempting to offer a definition or analysis of what a psychological or mental process or ability is, let the following incomplete list suffice to fix our ideas: perception, memory, imagination (classical Faculties); attention, motivation, consciousness, decision-making, problem-solving (processes or abilities that are the focus of much contemporary work in the cognitive sciences); and believing, desiring, intending, trying, willing, fearing, and hoping (common, folk psychological states).

This appears to be the standard approach within contemporary cognitive science and it is used to affirm the presence of cognition not only in individuals but also in groups.[28] Consider Cooke et al. (2013, 256) again:

> The term "cognition" used in the team context refers to cognitive processes or activities that occur at a team level. Like the cognitive processes of individuals, the cognitive processes of teams include learning, planning, reasoning, decision making, problem solving, remembering, designing, and assessing situations.

Accordingly, employing common-sense intuitions in order to judge whether some type of behavior may qualify as a cognitive process and whether, *eo ipso,* the underlying system is a

---

[27] For details on the debate on the 'mark of the cognitive', how it may be used against the hypotheses of extended and distributed cognition, and the considerable difficulty to come up with an unproblematic account for such a concept, see Clark (2010), Menary (2006), Adams and Aizawa (2001, 2008, 2010), Ross and Ladyman (2010) as well as Rupert (2011). On a different but related note, an anonymous referee points out that Huebner who supports, in Theiner's (forthcoming) terminology, the "computational stance" to group cognition would not be satisfied by the appeal to attitudinal behaviorism. Huebner additionally requires that the relevant cognitive task be performed on the basis of collective mental representations. However, the general dynamicist approach to cognition and the "dynamical stance" to group cognition (Theiner forthcoming) that the present approach falls under (see also fn. 22) avoid appealing to the indeterminate notion of mental representations, let alone to *collective* mental representations (for an overview on the debate of mental representations, as well as their relation to the "computational" and "dynamical stance", see (Pitt, 2013)). Appealing to mental representations therefore marks a fundamental methodological difference between the "dynamical" and the "computational stance" to cognition in general and group cognition in particular. As a side note, it is worth noting that cognitive scientists hardly ever appeal to the presence of mental representations in order to assert that a system qualifies as a cognitive system, precisely because there is no consensus (either within cognitive science or philosophy of mind) as to what mental representations are supposed to be.

[28] The Social Parity Principle (Theiner et al., 2010a; Theiner, 2013) puts forward essentially the same approach for recognising which group processes may count as cognitive.

cognitive system too, is not a form of philosophical behaviorism. Instead it is standard practice within contemporary cognitive science.

4. A DYNAMICAL APPROACH TO GROUP EMERGENCE

Let us assume that ADC is successful. Even so, taken on its own—in the absence of a plausible account of *downward causation*—it cannot provide sufficient support for the irreducibly *emergent* nature of group cognition and distributed cognitive systems. A failure to provide a clear understanding of how group cognitive entities can manifest causal effects over and above the sum of the causal powers exhibited by their individual members would simply amount to a failure to accommodate the problem raised by epiphenomenalism. Before concluding and in order to address this concern, it is important to revisit the central worries associated with emergence in light of the preceding discussion. This will bring to the fore a naturalistic understanding of downward causation that that has been already hinted at in the previous discussion.

Recall that, according to Kim (1984, 1989, 1993), if, at any one instance, there is a causal relation between the physical properties of two token events, it will be redundant to claim that some higher-level property (that supervenes on the physical properties of the realization base) is also causally responsible for the occurrence of the relevant events—specifically, the higher-level property will be epiphenomenal. To return to one of the examples we considered in §3.2, if, at some specific moment, the production and reception of certain visual cues and sounds between two human adults led to the recollection of a memory, then there is no need to claim that it is the group-level social property of transactive memory that led the TMS to effectively elicit the relevant piece of information. If in order to explain one particular instance of the correlation between two individual-level events one can point to individual-level properties alone, then there is no need to also posit additional group-level (social) properties and entities.

The previous discussion, however, can allow us to bypass the problem of epiphenomenalism while still respecting materialism. To start with the threat of epiphenomenalism, note that transactive memory may not only be multiply realized but its multiple realizability may be wildly disjunctive: [29] It may be the product of human beings interacting by using English, French, Chinese or even sign language. They may be in close proximity, or use Skype and, along with body language, they may even use written rather than spoken language. Even more wildly, we can imagine TMSs being instantiated by Martians with silicon brains, communicating not by talking to each other but by wirelessly transmitting information directly to each other's minds.

---

[29] For alternative discussions of the epiphenomenalist worry within the context of group cognition, see (Theiner & O'Connor, 2010, sections 2.2.3 and 4.1) and (Huebner, 2013, chapters 5-6).

In other words, the physical properties and linear relations of the realizing component parts of every instance of a TMS may have nothing in common (at least not from the point of view of physics, neurobiology or psychology). What all TMSs may share, instead, and which may allow them to manifest the properties and causal forces associated with them, are commonalities at a higher level of description. Specifically, the only thing that TMSs may be truly said to have in common in every one of their wildly realized instances, and which may be responsible for their distinctive causal forces, is a specific kind of dense mutual interactions between their realizing component parts. According to the 'systemic properties' and 'ongoing feedback loops' arguments, however, the dense mutual interactions from which these properties arise can only be defined by appealing to higher-level systems, whose properties and boundaries transcend those of the realizing component parts. In effect, higher-level systems like TMSs appear to have their distinctive place in the causal dynamics of the universe. Specifically, the causal explanation of the existence and the effects of properties like transactive memory, which are multiply realizable in a wild fashion, is possible only in terms of higher-level systems, such as TMSs.[30]

One can therefore insist on the necessity of postulating certain higher-level systems, because they do serious causal explanatory work: Granted, the existence of *some* (but no specific, due to multiple realizability) lower-level entities is indeed necessary for the higher-level properties to *possibly* arise; but in order for such properties (i.e., regular, or potentially regular, behaviors) to be actually manifested, what is further required is that the relevant dense mutual interactions between the contributing parts be in place, and thereby—according to DST—the corresponding higher-level entities too.

Now to see how all this is in line with materialism, we need to be clear about one thing: Nothing in the previous comments should be taken as denying that group-level social properties are material properties. Rather, the claim is that higher-level properties originate from a more complex level of materiality that cannot be fully captured by appealing to the lower-level properties of their component parts and/or their linear relations. In other words, the present approach recognizes that group-level social properties rely, first of all, on the individual-level material properties of their component parts and their linear relations—not every possible arrangements of individual-level properties can support the emergence of group cognition, after all. But once higher-level, group cognitive properties have sprung into

---

[30] This is not to say that all multiply realizable properties will lead to the postulation of higher-level entities. Following Fodor's (1974; 1997) rationale, if there are only a few realizing states, or if those states display some common features, the reduction of the higher-level properties to lower-level ones may still be performed unproblematically. If, however, the several possible underlying bases of a higher level property are an otherwise unrelated combination of many underlying concepts and terms (as is the case of properties that are both multiply *and* wildly realizable), then postulating the higher-level systems will be necessary for the reason explained above. Conversely, not all properties of every dynamical system are going to be multiply realizable. Whether this is going to be the case or not will each time depend on how easy it is for the parameter space of the target system to exceed bifurcation points. When small changes in the parameter space of a system are likely to cause bifurcations in its state space, the system will be less likely to be multiply realizable.

existence, they can have distinct, additional effects, over and above the individual-level properties of their underlying components. Yet, such additional effects—that go beyond the causal powers of the underlying individual components—are *material* effects all the same.

To elaborate, in order for any token event that may be potentially classified as the manifestation of a group property (take, for example, one instance of all the possible realization bases of transactive memory) to occur, the material properties of *some* physical, individual-level realization base and their linear relations must first be in place. Taken on its own, however, this is not sufficient for any token event to qualify as the manifestation of a group property. What is further required is that it be an instance of *regular (or potentially regular) behavior*. And for this to be the case, as opposed to merely being a fleeting, individual-level event, it must be an instance of the complex, non-linear—yet still *material*—component interactions that one of the higher-level, group properties is identified with.[31] In DST terminology, the event must be represented by one of the *collective variables* of the higher-level system and its regularity must be portrayed by the *limit sets* that shape the *trajectory* of the relevant *collective variable*.

In other words, *material* properties of both the group-level, social kind and the lower level, individual kind appear to be necessary and jointly sufficient for the occurrence of group-level events. In effect, this observation provides the means to bypass Kim's 'explanatory causal exclusion' principle, according to which "no event can be given more than one complete and independent explanation" (Kim 1989, p. 79). While this principle is probably correct, it does not generate any problems for the above analysis, as it recognizes that in order for an event to count as a group-level *property,* we need both an individual-level and a group-level explanation. In other words, the 'explanatory causal exclusion' principle creates no problems in cases such as the above, because in order to explain the occurrence of a collective property, neither the lower-level individual explanation nor the group-level social explanation are complete on their own. Instead, they may only count as *jointly sufficient*.

With respect to the previous example, this means that it is insufficient to claim that the production and reception of visual cues and sounds between two adult human beings is causally responsible for the ability to recollect a shared memory. The reason is that it is highly unlikely that the same physical exchanges between two different individuals or between the same individuals but under different circumstances are going to have the same positive effect. What leads to the successful and regular invoking of shared memories between individuals, such that they can exhibit the property of transactive memory, is not any of the specific underlying physical or individual-level processes but the individual's non-linear transactive

---

[31] An alternative way to put the idea is to express it in the following two steps: (a) individual-level properties and linear relations are necessary 'enabling' conditions; (b) once group level properties are in place, due to non-linear interactions between the individual members, they (group properties) can have distinct downward-causal effects on the individual members of the group.

communication processes. Even though there must indeed be some underlying physical and individual-level processes that will allow the transactive communication to be instantiated, what is distinctive in every case where the property of transactive memory is manifested is the higher-level, complex, non-linear communication processes that cannot be captured by either the language of physics or any individual-level scientific description. Yet, these higher, group-level processes are no less material than the underlying physical and individual activity.

According to the above then, both lower- and higher- level material properties are required for the manifestation of group cognition. The twist in the story, however, is that the lower-level physical and individual-level properties are not necessary in the same way the higher-level collective properties are, since the latter can be multiply realized. The existence of *one of* the appropriate realization bases will be necessary, but since there can be a multitude of them, no specific realization base is essential. In contrast, it is the existence of very specific complex interactions that will be crucial and it is in this sense that the relevant non-linear interactions provide the emerging properties and entities with their *identity*.

There is a possible worry, however: Could it be objected that the above goes against the principle of the 'causal closure of the physical', which states that all physical effects must have sufficient physical causes (Stoljar 2015; Kim 1997)? Strictly speaking, if we focus on the *terms* involved thereof, it does. Nevertheless, the preceding is well in line with the *spirit* of what the principle of the causal closure of the physical is meant to convey: Even though group-level, and in general all higher-level properties beyond the microphysical (and chemical) domain, are indeed not *physical* properties to be found in *any law of physics*, they still are *material* properties (even if material at a higher level of complexity). The fact that, in the context of the present topic, such material properties are called 'social' has only to do with the fact that they are associated with social groups. Nevertheless, there is absolutely no necessity to further associate such sociological claims with either substance or property dualist considerations: Group-level social properties of this kind just refer to material properties within a higher-level (materialist) science of sociology.

So to avoid any further future confusion it would perhaps be preferable to rephrase the principle of the 'causal closure of the physical' this way:

> Causal Closure of the *Material*: All effects must have a sufficient *material* cause.

How is this formulation different to the previous one? It accentuates the fact that some causal regular effects that are associated with higher-level (biological, psychological and sociological) properties cannot be captured by the language of physics. In other words, not all material causes are physical causes. It is this last claim that was erroneously implied by the previous version of the principle, and against which the new formulation is meant to act as a guard.

But if the above is true then the formulation of the *supervenience thesis* must be similarly amended too:

> *Material Supervenience*
>
> If two events share all of their material (as opposed to just physical) properties, they will share all of their social (or biological, or mental) properties.

By reformulating supervenience this way it becomes obvious again that the idea of sociological emergence does not really run against Kim's 'explanatory causal exclusion' principle:[32] It is possible to claim that two events are identical from the point of view of physics (or biology, or psychology) but not from the point of view of sociology, while still insisting that all events are material events (material properties of both the group-level, social kind and the lower-level, physical, biological and psychological kinds are necessary and jointly sufficient for the occurrence of group-level social events). Conversely, given multiple realizability, two events might be identical from the point of view of sociology but not of physics, biology or psychology, not because they share any non-material properties, but because their shared material properties (on the basis of which we can recognize them as a single sociological type of event) are material properties at a level of complexity that the language of the lower-level scientific descriptions cannot capture.

This inability of physics (and the rest of the natural sciences) to capture every complex material interaction, while also erroneously presupposing that all causes must be captured in terms of physics (or in terms that can be in principle reduced to the language of physics), has so far given the impression that downward causation must somehow arise *ex nihilo*, making the claim that there are additional properties to the properties that physics recognizes look particularly suspicious.[33] Given the present analysis, however, all causation, including downward causation, is in fact material causation. It just so happens that the essence of some regularities in the natural world does not depend on the properties that physics is concerned with, *even though it is determined by them.* The reason is that the properties described by physics allow for a multitude of ways in which matter may behave. On certain occasions, matter self-organises into more complex systems that restrain its behaviour. This behaviour is consistent with (i.e., determined by) all the possible physical (and in general, lower-level) behaviours of matter, but its *regularity* (such that it can be considered as a property in its own right) depends on the *coordinated* action of the parts of the overall system. In other words, it is

---

[32] It should be here noted that the above principle refers to local, rather than global supervenience. That is, the physical (biological and psychological) description of two group *entities* might be identical without them being sociologically identical. Nevertheless, if, in addition to their physical (biological and psychological) properties, two group entities also share the same sociological (yet still material) properties, they will also be sociologically identical. This is a form of local supervenience, because the sociological properties that determine whether the relevant group may qualify as a group entity in its own right are properties whose occurrence or absence depends only on the (non-linear) interactions of the components of the relevant group and no other external (global) factors constitutively affect their manifestation.

[33] It is for this same reason that, in §1, it was important to draw the subtle distinction between 'physicalism' and 'materialism'. The difference is that, according to materialsm, all properties are, or supervene on, *material*—as opposed to specifically physical—properties. See also fn. 3.

behaviour that occurs as part of the *collective variables* of the emergent system and thus belongs to it as a whole.

This is important, because, by understanding downward causation this way,[34] we no more have to be suspicious of cognitive scientists, who like Theiner et al. (2010), claim that "groups have the potential to display emergent cognitive properties that no individual member has, or might be capable of having." In principle, there could be several mere aggregates of individuals that could momentarily, in a fleeting way, exhibit behavior that resembles the properties of a distributed cognitive system. But in order for such behavior to be *regular* such that it can count as the property of a system, there must be a structure that can support, and at the same time be sustained by, the non-linear interactions that give rise to the relevant property. In other words, the higher-level system must be in place.

To close this section, then, it is helpful to summarize the above by noting the difference between reductionism and the present dynamical approach to emergence. Reductionism is the position that all phenomena are in principle reducible to the laws of physics. This is a problematic position because it identifies the physical world with the much broader material world, and thereby overestimates the explanatory power of physics. On the contrary, the present dynamical approach to emergence recognizes that all phenomena must be consistent with the laws of physics, but it denies that all phenomena can be reduced to physics. Instead, there are complex material properties, entities, and causal relations that can only be captured by higher-level systems of analysis, which postulate entities and properties that can constrain the behavior of the underlying physical basis just as much as they can be constrained by it.

5. CONCLUSION

In summary, the focus has been on the dynamic emergence of distributed cognitive systems. Following DST, the central claim is that the coupling of two or more elements on the basis of non-linear relations that arise out of ongoing mutual interactions provides a clear verdict for the manifestation of emergent properties and entities—properties and entities that exert downward causation by constraining their constituent parts to only just a few of the possible behaviors they would exhibit were they to act independently of each other. Moreover, this approach to sociological emergence is naturalistically respectable. It admits of only one form

---

[34] It is worth pointing out that the present approach to downward causation is not so different from Craver and Bechtel's (2007) approach to top-down causation as *constitution*. Craver and Bechtel argue that top-down causation is the restraints of mechanisms on their component parts. In the absence of the parts, there would be no overall system to constrain their subsequent behavior. This means that there is a symmetrical relationship between parts and the mechanisms they give rise to. Craver and Bechtel further note, however, that causal relationships have been traditionally thought of as *a*symmetrical relations. Top-down causation, which is symmetrical, should therefore be understood in terms of constitution rather than in causal terms.

of substance—i.e., matter—and even though it propounds the existence of several *types* of properties (i.e., physical, biological, individual as well as sociological ones) it insists that they all form proper subsets of the general category of material properties. To insist on the emergent status of group-level social properties and entities does not mean that they are fundamentally different from the relevant underlying properties and entities in any radically metaphysical sense. Rather, the reality of emergent sociological entities is different only in the following way: It arises out of a level of complexity that even though it is not captured by the properties and linear relations of the subvenient physical, biological and psychological entities, it imposes additional—yet still material—constraints on them.

ACKNOWLEDGEMENTS

I am thankful to Adam Carter for comments on an early draft of the paper. I am also thankful to two anonymous referees for *Minds and Machines.* This paper was produced as part of the AHRC-funded 'Extended Knowledge' research project (AH/J011908/1), which was hosted at Edinburgh's *Eidyn* Research Centre.

REFERENCES

Abraham, F. S., Abraham, R. H., & Shaw, C. (1990). A Visual Introduction to Dynamical Systems Theory for Psychology. Santa Cruz, CA: Aerial Pr.

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

Adams, F., & Aizawa, K. (2008). *The bounds of cognition*. Blackwell Publishing Ltd.

Adams, F., & Aizawa, K. (2010). Defending the bounds of cognition. In Menary (Ed.), *The extended mind*. Cambridge, Massachusetts: MIT Press.

Attanasi, A., Cavagna, A., Del Castello, L., Giardina, I., Jelic, A., Melillo, S. & Viale, M. (2015). Emergence of collective changes in travel direction of starling flocks from individual birds' fluctuations. *Journal of The Royal Society Interface*, *12*(108), 20150319.

Arrow, H., McGrath, J., Berdahl, J. (2000). Small Groups as Complex Systems: Formation, Coordination, Development, and Adaptation. Sage Publications.

Barnier, A. J., Sutton, J., Harris, C. B., & Wilson, R. A. (2008). A conceptual and empirical framework for the social distribution of cognition: The case of memory. Cognitive Systems Research, 9(1–2), 33–51. doi:10.1016/j.cogsys.2007.07.002

Becco, C., Vandewalle, N., Delcourt, J., & Poncin, P. (2006). Experimental evidences of a structural and dynamical transition in fish school. *Physica A: Statistical Mechanics and its Applications*, *367*, 487-493.

Beckermann, A., Flohr, H., & Kim, J. (1992). Emergence Or Reduction?: Essays on the Prospects of Nonreductive Physicalism. Walter de Gruyter.

Beer, R. D. (1995). A dynamical systems perspective on agent-environment interaction. Artificial Intelligence, 72(1–2), 173–215. doi:10.1016/0004-3702(94)00005-L

Bickle, J. (2013). Multiple Realizability. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Spring 2013.). Retrieved from http://plato.stanford.edu/archives/spr2013/entries/multiple-realizability/

Bonabeau, E., & Meyer, C. (2001). Swarm intelligence: A whole new way to think about business. *Harvard business review*, *79*(5), 106-115.

Bressler, S. L., & Kelso, J. A. S. (2001). Cortical coordination dynamics and cognition. Trends in Cognitive Sciences, 5(1), 26–36. doi:10.1016/S1364-6613(00)01564-3

Campbell, Donald T. (1974). “Downward causation in hierarchically organised biological systems,” in Francisco Jose Ayala and Theodosius Dobzhansky, eds., *Studies in the Philosophy of Biology: Reduction and Related Problems*, London/Basingstoke: Macmillan.

Campbell, R. J., & Bickhard, M. H. (2011). Physicalism, emergence and downward causation. *Axiomathes,* 21(1), 33-56.

Chemero, A. (2009). Radical embodied cognitive science. Cambridge, Mass.: MIT Press.

Clark, A. (2010). Coupling, constitution, and the cognitive kind: A reply to Adams and Aizawa. In Menary (Ed.), *The extended mind*. Cambridge, Massachusetts: MIT Press.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive science*, *37*(2), 255-285.

Coey, C. A., Varlet, M., & Richardson, M. J. (2012). Coordination dynamics in a socially situated nervous system. *Frontiers in human neuroscience*, *6*.

Corradini, A., & O'Connor, T. (Eds.). (2010). Emergence in science and philosophy. Routledge.

Craver, C. F., & Bechtel, W. (2007). Top-down Causation Without Top-down Causes. Biology & Philosophy, 22(4), 547–563. doi:10.1007/s10539-006-9028-8

Dale, R., & Spivey, M. J. (2006). Unraveling the Dyad: Using Recurrence Analysis to Explore Patterns of Syntactic Coordination Between Children and Caregivers in Conversation. Language Learning, 56(3), 391–430. doi:10.1111/j.1467-9922.2006.00372.x

Dale, R., Fusaroli, R., Duran, N., & Richardson, D. C. (2013). The self-organization of human interaction. *Psychology of learning and motivation*, *59*, 43-95.

Pitt, David, "Mental Representation", The Stanford Encyclopedia of Philosophy (Fall 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2013/entries/mental-representation/>.

Davidson, D. (1995). Laws and cause*. *Dialectica,* 49(2-4), 263–280. doi:10.1111/j.1746-8361.1995.tb00165.x

Davidson, D. (2002). Essays on Actions and Events: Philosophical Essays Volume 1 (New Ed edition.). Oxford : New York: Clarendon Press.

Dennett, D. C. (1993). Consciousness Explained (New Ed edition.). London: Penguin.

Duarte, R., Araújo, D., Folgado, H., Esteves, P., Marques, P., & Davids, K. (2013*a*). Capturing complex, non-linear team behaviours during competitive football performance. Journal of Systems Science and Complexity, 26(1), 62-72.

Duarte, R., Araújo, D., Correia, V., Davids, K., Marques, P., & Richardson, M. J. (2013*b*). Competing together: Assessing the dynamics of team–team and player–team synchrony in professional association football. *Human movement science*, *32*(4), 555-566.

Duarte, R., Araújo, D., Correia, V., & Davids, K. (2012). Sports teams as superorganisms. *Sports medicine*, *42*(8), 633-642.

Emmeche, C., Køppe, S., & Stjernfelt, F. (2000). Levels, emergence, and three versions of downward causation. In Peter Bøgh Andersen, Claus Emmeche, Niels Ole Finnemann and Peder Voetmann Christiansen, eds., Downward causation. Minds, bodies and matter, 13-34.

Fodor, J. (1997). Special Sciences: Still Autonomous After All These Years. Noûs, 31, 149–163.

Fodor, J. A. (1974). Special Sciences (Or: The Disunity of Science as a Working Hypothesis). Synthese, 28(2), 97–115.

Froese, T., Gershenson, C., & Rosenblueth, D. A. (2013). The Dynamically Extended Mind -- A Minimal Modeling Case Study. arXiv:1305.1958 [nlin]. Retrieved from http://arxiv.org/abs/1305.1958

Fusaroli, R. Rączaszek-Leonardi, J. Tylén, K.(2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*, 147-157.

Fusaroli, R., Gangopadhyay, N., & Tylén, K. (2014). The dialogically extended mind: Language as skilful intersubjective engagement. *Cognitive Systems Research*, *29*, 31-39.

Fusaroli, R., & Tylén, K. (2013). Linguistic coordination: Models, dynamics and effects.

Fusaroli, R., & Tylén, K. (2015). Investigating conversational dynamics: Interactive alignment, Interpersonal synergy, and collective task performance. *Cognitive science*.

Gelder, T. van. (1995). What Might Cognition Be If Not Computation? Journal of Philosophy, 92(7), 345–81.

Giere, R. (2002*a*). 'Discussion Note: Distributed Cognition in Epistemic Cultures'. *Philosophy of Science*, 69.

—— (2002*b*). 'Scientific Cognition as Distributed Cognition'. In *Cognitive Bases of Science*, eds. Peter Carruthers, Stephen Stitch and Michael Siegal, Cambridge: Cambridge University Press, 2002.

Graham, George, Behaviorism, *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2015/entries/behaviorism/>.

Harris, C. (2010). Collaborative Remembering: When Can Remembering With Others Be Beneficial? (pp. 131–134). Macquarie Centre for Cognitive Science. doi:10.5096/ASCS200921

Heylighen, F., Heath, M., & Van, F. (2004). The Emergence of Distributed Cognition: a conceptual framework. In Proceedings of Collective Intentionality IV.

Hollingshead, A. B. (1998a). Communication, Learning, and Retrieval in Transactive Memory Systems. Journal of Experimental Social Psychology, 34(5), 423–442. doi:10.1006/jesp.1998.1358

Hollingshead, A. B. (1998b). Retrieval processes in transactive memory systems. Journal of Personality and Social Psychology, 74(3), 659–671. doi:10.1037/0022-3514.74.3.659

Hollingshead, A. B., & Brandon, D. P. (2003). Potential Benefits of Communication in Transactive Memory Systems. Human Communication Research, 29(4), 607–615. doi:10.1111/j.1468-2958.2003.tb00859.x

Huebner, B. (2013). Macrocognition: Distributed minds and collective intentionality. New York: Oxford University Press.

Humphreys, Paul (1997). "How Properties Emerge," *Philosophy of Science*, 64: 1–17.

—— "Emergence, Not Supervenience," *Philosophy of Science*, 64: S337-S345.

Hutchins, E. (1996). Cognition in the Wild (New edition edition.). Cambridge, Mass.: MIT Press.

Kelso, J. A. S., & Engstrøm, D. A. (2008). The Complementary Nature. Cambridge, Mass.; London: A Bradford Book.

Kelso, J. S. (1997). *Dynamic patterns: The self-organization of brain and behavior*. MIT press.

Kim, J. (1984). Epiphenomenal and Supervenient Causation. *Midwest Studies in Philosophy*, 9(1), 257–70.

Kim, J. (1989). Mechanism, Purpose, and Explanatory Exclusion. *Philosophical Perspectives*, 3, 77–108. doi:10.2307/2214264

Kim, J. (1993). Supervenience and Mind, Cambridge: Cambridge University Press.

Kim, J. (1997). Does the problem of mental causation generalize?. In *Proceedings of the Aristotelian Society*, pp. 281-297.

Kim, J. (1999). Making Sense of Emergence. *Philosophical Studies*, 95(1-2), 3–36. doi:10.1023/A:1004563122154

Knorr-Cetina, K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge*. Harvard University Press.

Lewis, K. (2003). Measuring transactive memory systems in the field: scale development and validation. *The Journal of Applied Psychology*, 88(4), 587–604.

Ludwig, K. (2015). Is Distributed Cognition Group Level Cognition?, *Journal of Social Ontology*. Volume 1, Issue 2, Pages 189–224

Li, L., Peng, H., Kurths, J., Yang, Y., Schellnhuber, H.J. (2014): Chaos-order transition in foraging behavior of ants. *Proceedings of the National Academy of Sciences,* Early Edition [DOI:10.1073/pnas.1407083111]

Li, L., Yang, Y., & Peng, H. (2009). Fuzzy system identification via chaotic ant swarm. *Chaos, Solitons & Fractals*, *41*(1), 401-409.

Luisi, P. L. (2002). Emergence in Chemistry: Chemistry as the Embodiment of Emergence. Foundations of Chemistry, 4(3), 183–200. doi:10.1023/A:1020672005348

Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science*, *1*(2), 320-339.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). *The appeal of parallel distributed processing* (pp. 3-44). Cambridge, MA: MIT Press.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in cognitive sciences*, *14*(8), 348-356.

McGrath, J. E., Arrow, H., & Berdahl, J. L. (2000). The Study of Groups: Past, Present, and Future. *Personality and Social Psychology Review*, 4(1), 95–105. doi:10.1207/S15327957PSPR0401_8

Menary, R. (2006). Attacking the bounds of cognition. *Philosophical Psychology*, 19(3), 329–344.

Minsky, M. (1988). The Society of Mind (Pages Bent edition.). New York: Pocket Books.

Moreland, R. (1999). Transactive memory: Learning who knows what in work groups and organizations. In L. Thompson, J. Levine, & D. Messick (Eds.), (pp. 3–31). Lawrence Erlbaum Associates Publishers.

Morganti, M. (2009). A New Look at Relational Holism in Quantum Mechanics. Philosophy of Science, 76(5), 1027–1038.

Murphy, N., Ellis, G., & O'Connor, T. (Eds.). (2009). *Downward causation and the neurobiology of free will.* Springer Science & Business Media.

Nersessian, N. J. (2006). The Cognitive-Cultural Systems of the Research Laboratory. Organization Studies, 27(1), pp. 125-145

Niwa, H. S. (1994). Self-organizing dynamic model of fish schooling. *Journal of theoretical Biology*, *171*(2), 123-136.

Obuko, A. (1986). Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. *Advances in biophysics*, *22*, 1-94.

O'Connor, T., & Wong, H. Y. (2012). Emergent Properties. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Spring 2012.). Retrieved from http://plato.stanford.edu/archives/spr2012/entries/properties-emergent/

O'Connor, Timothy (1994). "Emergent Properties," *American Philosophical Quarterly*, 31: 91–104.

Palermos, S. O. (2014). Loops, constitution, and cognitive extension.*Cognitive systems research*, *27*, 25-41.

Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas.* New York, NY, USA: Basic Books, Inc.

Parunak, H. V. D. (1997). " Go to the ant": Engineering principles from natural multi-agent systems. *Annals of Operations Research*, *75*, 69-101.

Peng, H., Li, L., Yang, Y., & Liu, F. (2010). Parameter estimation of dynamical systems via a chaotic ant swarm. *Physical Review E*, *81*(1), 016207.

Port, R. F., & Van Gelder, T. (1995). *Mind as motion: Explorations in the dynamics of cognition*. MIT press.

Raczaszek-Leonardi, J., & Kelso, J. A. S. (2008). Reconciling symbolic and dynamic aspects of language. *New Ideas in Psychology*, 26(2), 193–207. doi:10.1016/j.newideapsych.2007.07.003

Richardson, M., Dale, R., & March, L. (2014). Complex Dynamical Systems in Social and Personality Psychology. *Handbook of research methods in social and personality psychology*, 253.

Riley, M. A., Richardson, M. J., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in psychology*, *2*, 38.

Rodriguez, E., George, N., Lachaux, J.-P., Martinerie, J., Renault, B., & Varela, F. J. (1999). Perception's shadow: long-distance synchronization of human brain activity. *Nature*, 397(6718), 430–433. doi:10.1038/17120

Ross, D., & Ladyman, J. (2010). The alleged coupling-constitution fallacy and the mature sciences. In Menary (Ed.), *The extended mind*. Cambridge, Massachusetts: MIT Press.

Ruben, D.-H. (1985). *The Metaphysics of the social world*. London: Routledge.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. (1986). Sequential thought processes in PDP models. *V*, *2*, 3-57.

Rupert, R. (2005). Minding one's cognitive systems: When does a group of minds constitute a single cognitive unit? *Episteme: A Journal of Social Epistemology* 1, 177–88.

Rupert, R. (forthcoing*a*). Individual Minds as Groups, Group Minds as Individuals

(University of Colorado, Boulder). In B. Kaldis (Ed.), *Mind and Society: Cognitive Science Meets the Philosophy of the Social Sciences*, Synthese Library Special Volume.

Rupert, R. (forthcoming*b*). Against group cognitive states. In S. Chant, F.

Hindriks, and G. Preyer (eds.), *From Individual to Collective Intentionality* (Oxford: Oxford University Press).

Rupert, R. D. (2011). Empirical arguments for group minds: A critical appraisal. *Philosophy Compass*, *6*(9), 630-639.

Sawyer, R. K. (2001). Emergence in Sociology: Contemporary Philosophy of Mind and Some Implications for Sociological Theory. *American Journal of Sociology,* 107(3), 551–585. doi:10.1086/338780

Sawyer, R. K. (2002). Nonreductive Individualism Part I—Supervenience and Wild Disjunction. *Philosophy of the Social Sciences*, 32(4), 537–559. doi:10.1177/004839302237836

Sawyer, R. K. (2003). Nonreductive Individualism Part II—Social Causation. *Philosophy of the Social Sciences*, 33(2), 203–224. doi:10.1177/0048393103033002003

Schmidt, R. C., Bienvenu, M., Fitzpatrick, P. A., & Amazeen, P. G. (1998). A comparison of intra-and interpersonal interlimb coordination: coordination breakdowns and coupling strength. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 884.

Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In *Coordination: Neural, behavioral and social dynamics* (pp. 281-308). Springer Berlin Heidelberg.

Sellars, W., 1963. "Philosophy and the Scientic Image of Man", in *Science, Perception, and Reality*, New York: Routledge & Kegan Paul, pp. 1–40.

Spivey, M. J., & Dale, R. (2006). Continuous Dynamics in Real-Time Cognition. *Current Directions in Psychological Science*, 15(5), 207–211. doi:10.1111/j.1467-8721.2006.00437.x

Spivey, M. (2007). *The Continuity of Mind.* Oxford University Press.

Stephan, A. (1999). Varieties of Emergence. *Evolution and cognition*, *5*(1), 50-59.

Stephan, A. (2006). The dual role of 'emergence'in the philosophy of mind and in cognitive science. *Synthese*, *151*(3), 485-498.

Stoljar, D. (2015). Physicalism. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy. Retrieved from http://plato.stanford.edu/archives/fall2009/entries/physicalism/

Sutton, J. (2008). Between Individual and Collective Memory: Interaction, Coordination, Distribution. *Social Research*, 75(1), 23–48.

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. doi:10.1007/s11097-010-9182-y

Teller, P. (1986). Relational Holism and Quantum Mechanics. *British Journal for the Philosophy of Science*, 37(1), 71–81.

Theiner, G. (2013a). Onwards and Upwards with the Extended Mind: From Individual to Collective Epistemic Action. In L. Caporael, J. Griesemer, & W. Wimsatt (Eds.), *Developing Scaffolds* (pp. 191–208). MIT Press.

Theiner, G. (2013b). Transactive Memory Systems: A Mechanistic Analysis of Emergent Group Memory. *Review of Philosophy and Psychology*, 4(1), 65–89. doi:10.1007/s13164-012-0128-x

Theiner, G. (forthcoming). Group-Sized Distributed Cognitive Systems. In Ludwig, K. & Jankovic, M. (Eds), *The Routledge Handbook of Collective Intentionality,* New York Routledge.

Theiner, G., Allen, C., & Goldstone, R. L. (2010). Recognizing group cognition. *Cognitive Systems Research*, 11(4), 378–395. doi:10.1016/j.cogsys.2010.07.002

Theiner, G., & O'Connor, T. The Emergence of Group Cognition. In Corradini, A., & O'Connor, T. (Eds.), *Emergence in science and philosophy*. Routledge.

Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. MIT press.

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

Tollefsen, D., & Dale, R. (2012). Naturalizing joint action: A process-based approach. *Philosophical Psychology*, 25(3), 385–407. doi:10.1080/09515089.2011.579418

Tollefsen, D. P. (2015). *Groups as agents.* John Wiley & Sons.

Tollefsen, D. P. (2006). From extended mind to collective mind. *Cognitive Systems Research*, 7(2–3), 140–150. doi:10.1016/j.cogsys.2006.01.001

Turnstrøm, K., Katz, Y., Ioannou, C. C., Huepe, C., Lutz, M. J., & Couzin, I. D. (2013). Collective states, multistability and transitional behavior in schooling fish. *PLoS Comput Biol*, *9*(2), e1002915.

Turvey, M. T. (1990). Coordination. *American psychologist*, *45*(8), 938.

Tylén, K., Fusaroli, R., Bundgaard, P. F., & Østergaard, S. (2013). Making sense together: A dynamical account of linguistic meaning-making. *Semiotica*, *2013*(194), 39-62.

Varela, F. J. (1993). The Embodied Mind: Cognitive Science and Human Experience (New edition edition.). Cambridge, Mass.: MIT Press.

Varela, F. J., & Singer, W. (1987). Neuronal dynamics in the visual corticothalamic pathway revealed through binocular rivalry. *Experimental Brain Research*, 66(1), 10–20.

Varela, F., Lachaux, J.-P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2(4), 229–239. doi:10.1038/35067550

Wegner. (1986). *Theories of group behavior*. New York: Springer-Verlag.

Wegner, D. M., Giuliano, T., & Hertel, P. T. (1985). Cognitive Interdependence in Close Relationships. In D. W. Ickes (Ed.), *Compatible and Incompatible Relationships* (pp. 253–276). Springer New York. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4612-5044-9_12

Wegner, Erber, R., & Raymond, P. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology,* 61, 923–929.

Wegner, D. M. (1995). A computer network model of human transactive memory. *Social Cognition*, 13, 319–339.

Warren, W. H., & Fajen, B. R. (2004). Behavioral dynamics of human locomotion. *Ecological Psychology*, *16*(1), 61-66.

Warren, W. H. (2006). The dynamics of perception and action. *Psychological review*, *113*(2), 358.

Wilson, J. (forthcoming). Metaphysical Emergence: Weak and Strong. In Tomasz Bigaj & Christian Wuthrich (eds.), *Metaphysics in Contemporary Physics*. Poznan Studies in the Philosophy of the Sciences and the Humanities.

Wilson, J. (2013). Nonlinearity and Metaphysical Emergence. In Stephen Mumford & Matthew Tugby (eds.), *Metaphysics and Science*.

Wilson, R. A. (2005). Collective memory, group minds, and the extended mind thesis. *Cognitive Processing,* 6(4), 227–236. doi:10.1007/s10339-005-0012-z

Wilson, R, A. (2001). Group-Level Cognition. *Philosophy of Science,* 68 (3), pp. 262-273.

Wimsatt, W. C. (1986). Forms of aggregativity. In M. G. Grene, A. Donagan, A. N. Perovich, & M. V. Wedin (Eds.), *Human Nature and Natural Knowledge* (pp. 259–291). Dordrecht: Reidel.

Wimsatt, W. C. (2000). Emergence as Non-Aggregativity and the Biases of Reductionisms. Foundations of Science, 5(3), 269–297. doi:10.1023/A:1011342202830