A Framework for Thinking About Distributed Cognition

Pierre Poirier
Department of philosophy and Cognitive Sciences Institute

Guillaume Chicoisne
Cognitive Sciences Institute

Université du Québec à Montréal

Abstract

As is often the case when scientific or engineering fields emerge, new concepts are forged or old ones are adapted. When this happens, various arguments rage over what ultimately turns out to be conceptual misunderstandings. At that critical time, there is a need for an explicit reflection on the meaning of the concepts that define the field. In this position paper, we aim to provide a reasoned framework in which to think about various issues in the field of distributed cognition. We argue that both relevant concepts, distribution and cognition, must be understood as continuous. As it is used in the context of distributed cognition, the concept of distribution is essentially fuzzy, and we will link it to the notion of emergence of system-level properties. The concept of cognition must also be seen as fuzzy, but for different a reason: due its origin as an anthropocentric concept, no one has a clear handle on its meaning in a distributed setting. As the proposed framework forms a space, we then explore its geography and (re)visit famous landmarks.

1. Introduction

Are companies, organisations and communities into the cognitive business just as we are? If you do your writing on OpenOffice's text editor, you are using an artefact for the conception of which a loosely connected community of individuals, spread throughout the world, have collaborated towards the common goal of providing free, open-source software to individuals worldwide. Can that community be said to be, in some sense, a collective cognitive system, manifesting distributed cognition, whose collective behaviour was the construction of an artefact no individual human programmer could ever construct by himself? Perhaps. But consider this: How is what they do different, really, from the process that constructed MS-WORD? Apart from the surely contingent fact that Microsoft's programmers and engineers were hired and paid by the corporation, is there any relevant cognitive difference between the community of people that Microsoft hires to write its product and the community that selflessly collaborate to produce OpenSource software? Is Microsoft a cognitive system? What about Ford or Toyota? And what about universities, news agencies, courts of law, parliaments, and military units? Take a court of law. Its purpose is explicitly cognitive: to determine, beyond reasonable doubt, the truth-value of certain propositions (X is guilty of crime C). To do so, a number of individuals have specific roles to play, each of them cognitive. Police officers have to *investigate* and *find facts relevant to the case*; defence

attorneys have to *research* (the *declarative external memories* that are) the various codes (criminal, civil, etc.) and precedents repertories for laws and precedents that are relevant to their case and, based on those, both *interpret* and present the case in a manner that is most favourable to the accused and *rebut* the prosecution's arguments; prosecutors have to do the same but in a manner that is least favourable to the accused; the judge has *insure the validity of the whole procedure*; and finally the jury has *listen* to all the evidence presented to it, *evaluate* it, *deliberate*, and *render judgment* as to the truth or falsity of the propositions presented to it by the court. The system is so built that, when everything works well (which is not always the case), none of the beliefs of the individuals involved *determines* the outcome. The decision process is, in some sense, supra-individual. As a cognitive process, evaluating the truth of the propositions is a system-level affair. This is the difference between living in a society with a justice system versus living in a society with vigilante justice (in French: the difference between "un système de justice" and "un justicier "). Now, is this description just a loose and ultimately unproductive analogy with the human mind? Perhaps courts of laws were design with a eye on the only model of a truth-finding system that was at hand at the time, that is, the human mind. Or perhaps, historians say, the human mind was conceived (in its modern form) with an eye on the only truth-finding system known at the time, that is, courts of law. Either proposition would explain the analogy. But the question remains: Is this mere analogy or is there more substance to the claim that there can be supra-individual systems manifesting distributed cognition? Some are quite content to view companies as cognitive systems of sorts, possessing institutional memory (which, like ours, may be tacit or explicit, declarative or procedural), decision capacities, creativity, and so on (Nonaka and Tekeuchi 1995, Hutchins 1996). Is this just conceptual confusion or clever marketing hype?

The foregoing only focused on humans. But what about ant colonies or other collectives of eusocial animals? It is possible to view them as superorganism, that is, organisms made up of other organisms*;* or as Minsky described them: "Genetically, the swarms of social ants and bees are really multibodied individuals whose different organs move around freely" (Minsky 1985). And if this can be said from a physical standpoint, shouldn't cognitive scientists be ready to study the (distributed) cognitive properties of superorganisms? Some are already quite happy to, speaking of "swarm intelligence," "ant colony optimization" and the like. And what about the multi-agent systems built by engineers, which will come to play an increasingly important role in a high-bandwidth networked society? Are these people just confused or are they on to something?

To address these questions, two conceptual options are possible. The first, conservative, insists on the differences between supra-individual and human cognition to enforce a traditional and strict reading of the relevant concepts. The second, liberal, emphasizes the similarities between supra-individual and human cognition to re-organize the conceptual landscape of cognition. Both options, we believe, illegitimately draw their rhetorical strength by imposing sharp readings on two fuzzy[1] concepts: distribution and cognition. In the context of cognition, distribution, as we'll argue, is essentially fuzzy (see below). And in the context of distribution, cognition, as we'll also argue, must, at least for now, be given a family-resemblance definition, best captured by a multidimensional fuzzy concept. In this position paper, we propose a framework that restores the essentially vague nature of the relevant concepts and explore how various usual suspects (e.g., neural networks, ant colonies, multi-agent systems, human organizations) fare in this new conceptual landscape.

2. A notion of distributed cognition

Proponents and opponents of distributed cognition have a definite, though usually not clearly explicit, idea of what "distribution" means in the context of cognition. We believe this underlying idea is linked to the notion of emergence. Truly distributed cognition is emergent cognition. One truly has a distributed cognitive system when one has a system where a new cognitive property *emerge*s from the interaction between the system's components, which may themselves be cognitive systems. This link between distributed cognition and emergence is fine, we believe. The problem is that people usually work with an all or nothing conception of emergence inherited from the early 20th century. Emergentism is often caricatured as the thesis claiming that *the total is more that sum of its part*. Note that something either is or isn't more than the sum of its parts. *This* concept of emergence is binary (in logic: classical). Hence, viewing distributed cognition as emergent cognition, something either is or isn't a case of distributed cognition. But things need not be this way. We will ground this discussion on a notion of emergence (inspired by engineering) that makes emergence an essentially fuzzy concept.

W. Wimsatt, a philosopher of biology trained as an engineer, offers a notion of emergence that is perfectly consistent with the current ontology of science (as opposed to the form of emergentism that was popular, say, at the beginning of the 20th century). He defines emergence as a failure of aggregativity. Take a property P of a system S, s_1 to s_m being the *m* components of S, p_1 to p_n the *n* properties, $p_1(s_1)$, $p_2(s_1)$, …, $p_n(s_1)$, $p_1(s_2)$, $p_2(s_2)$, …, $p_n(s_2)$, … $p_1(s_m)$, $p_2(s_m)$, …, $p_n(s_m)$ properties of *S's* components, and the organisation or interaction mode *F* of these component properties. P of S may be defined thus:

- $P(S) = F[p_i(s_j)$ for i=1 to n and j=1 to m]

P(S) is aggregative to the extent four conditions are respected:

1. *Condition IS (invariance under substitution): P*(*S*) is invariant under intersubstitution of the parts of S with one another or under substitution of one or more of the parts with other parts from a domain of relevantly similar parts.
2. *Condition QS (qualitative invariance)*: *P*(*S*) remains qualitatively similar (differing only in value) under addition or subtraction of parts.
3. *Condition RA (reaggregation):* The composition function for *P*(*S*) is invariant under operations involving decomposition or reaggregation of parts.
4. *Condition CI (cooperation/inhibition)*: There are no cooperative or inhibitory interactions among the parts of the system. (Wimsatt 1986)

Take a pile of books. It is (in some sense) a system and has a number of properties, among them a certain mass. Consider then the property *Mass*(pile of books), or for short M(b). Now, you can interchange the position of the books in the pile or replace one book with one that is similarly relevant with respect to M(b), that is replace it with another book of the same mass (say I replace your copy of Kurt Goldstein's *The Organism* with mine). Clearly M(b) is invariant under such substitutions: their is no failure of condition IS. Also, books can be added or subtracted from the pile without the pile either losing its property of mass or mass becoming some qualitatively different property. M(b) is qualitatively invariant under

the operations of addition and subtraction of books: there is no failure of condition QS. Moreover, individual books from the pile may be decomposed (e.g., put trough a shredder) or reaggregated (good luck!) and M(b) will stay the same. M(b) is invariant under the operations of decomposition or reaggregation: there is no failure of condition RA. Finally, with respect to M(b) there are no relevant interaction (cooperative or inhibitory) between the books in the pile. There is no failure of condition CI. It follows that M(b) is a completely aggregative (hence non-emergent) property. Note on the other hand that the height of the pile, H(b), fails condition CI (there are inhibitory interactions between the books in the pile: different ways of organising the books in the pile (vertically, horizontally) gives different heights to the pile. Since H(b) does not fail any other conditions of aggregativity, we may say that H(b) is a mostly aggregative property (or a mildly emergent one).

As Wimsatt points out, viewing emergence as the failure of conditions of aggregativity has the advantage of seeing emergence as a continuum between fully aggregative properties and fully emergent properties, both of which are rare in nature. Most properties fall somewhere on this continuum, as determined by the extent they satisfy or fail aggregativity conditions 1 through 4. Note also, as Wimsatt points out, that since emergence, thus defined, presupposes the existence of a composition function (*F*), it is not the opposite of reducible. All systemic properties on the aggregativity/emergence continuum are in principle reducible to the properties of the system's component, their mode of organisation and their interactions. This is what's called "objective emergence": emergent properties are real properties, in principle reducible to lower level properties. Whether they can in practice be reduced depends on our knowledge of (1) the system's components and their properties and (2) the systemic interactions between them. If we know nothing or very little of either or if these are complex (relatively to human's limited computation abilities), then the more a systemic property is emergent the more it will seem to "appear" (almost magically) from nothing. This condition is called "subjective emergence". The important point about subjective emergence is that it is just that: subjective (i.e., an epistemic reflection of our ignorance).

Now that we have explained what we believe the concept of emergence underlying any theoretical reflection about distributed cognition should be, we must do the same with the concept of cognition. Why is this important? Doesn't everyone know what cognition is? In the present context, we need to be careful about that. The novel, but controversial, idea behind the notion of distributed cognition is that we may ascribe cognitive properties to types of systems that haven't been traditionally seen as the bearers of such properties: ant colonies, organisations, research labs, and so on. In this context, a number of things about the concept of cognition should be kept in mind. (1) We need to define cognition in a way that is not overly chauvinistic, for instance by tying cognition to a manifestly human manifestations of cognition. Cognitive systems may need sensors, but they do not need ours (including their felt qualities). Cognitive systems may make their living in an environement, but they do not need to make their living in our physical environment, even less in our own ecological niche. Cognitive systems may need warning signals that tells them something is going on that treathens there integrity, but they do not need our warning systems (pain, anxiety, fear, stress). Cognitive systems need to persist for some time, but they do not need to do so by being part of biological life on earth. In short, we need a conception of cognition that is not overly anthropocentric (or chauvinistic) and that can allow the ascription of cognitive properties to systems that are not typically viewed as such. Otherwise, the whole issue of whether distributed cognition is possible or not will be decided by conceptual

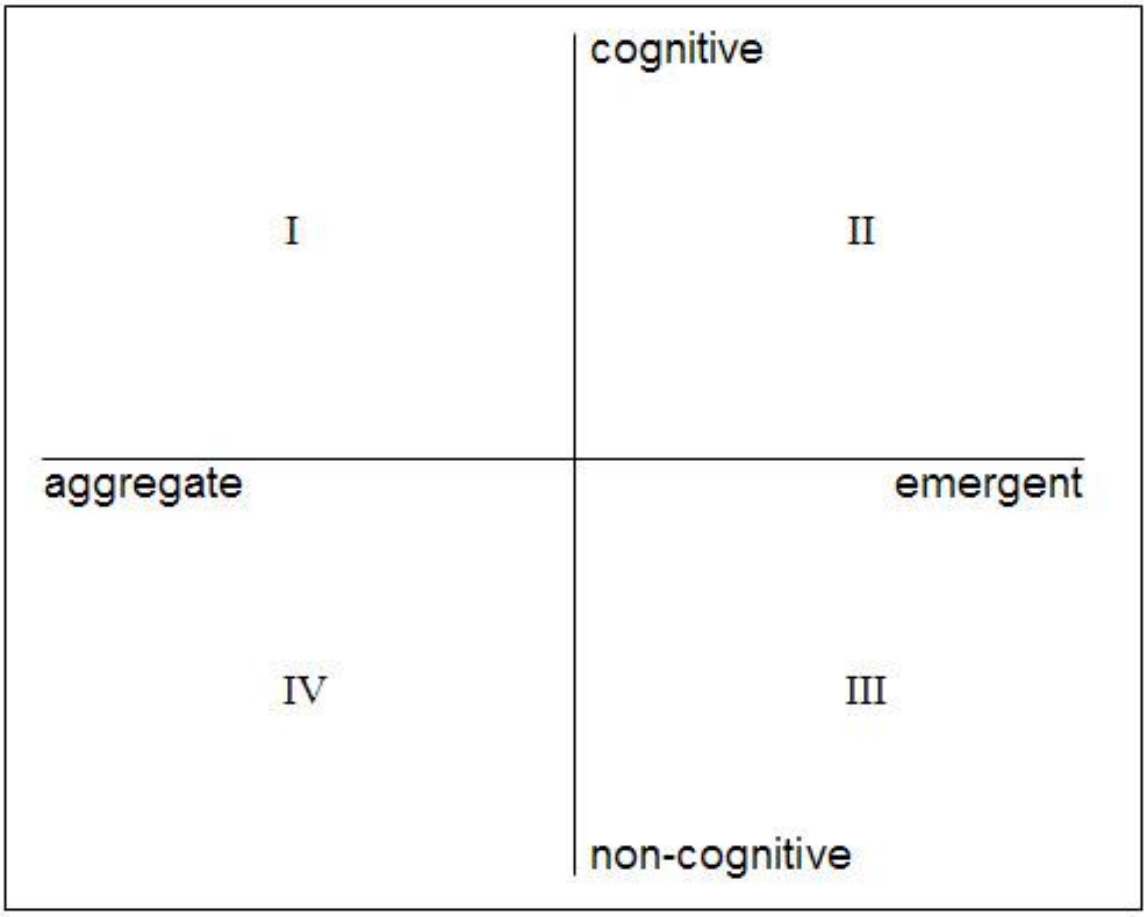
decree, that is by insisting on an anthropocentric concept of cognition that *makes* distributed cognition impossible. Once any debate over ideas has been reduced to that level, it usually becomes an unproductive fight over the meaning of words. (2) But the problem that concerns us goes deeper than any councel for caution against overly anthropocentric conceptions of cognition could solve. We have to be open to the possibility that the cognitive property in question may be quite alien to us and that, as a consequence, we may be literally blind to it. Think of it this way. Autists are, it is said (Baron-Cohen 1997), blind to some mental properties in others, that is, they are be mindblind. The question is: if some truly novel cognitive property emerges out of the interaction of agents, will we be as blind to it as autists are (according to Baron-Cohen) blind to our (plain old run of the mill) mind. This is related to the classic sci-fi plot-maker: will we recognize an alien intelligence staring us right in the face? We don't propose to resolve this old chestnut here (it may actually be impossible to do so). But it is important to keep a broad mind when thinking about cognition in the present context. (3) However, the very fact that what we are looking for may be quite alien to us should impose a measure of caution. No one wants to ascribe cognitive properties to Gaia, Jupiter's Red Spot or auto-catalytic chemical reactions. Talk of cognition in the context of distributed cognition invites analogy, metaphor, and loose-talk. While these may be necessary in science, especially when disciplines are young, as any cognitive scientist who ever bought into the "computer metaphor" should know, they are also dangerous tools that can quickly discredit any aspiring young field (think of memetics). (4) Cognitive properties should not be tied too closely together to avoid illegitimate inferences. No one wants to infer that ant colonies must be *conscious* since they can collectively optimize some parameter. Or, conversely, no one wants to deny ant colonies some cognitive properties because they are obviously not conscious. Bearing these points in mind, we propose to explicitly view cognition as a cluster concept, that is, a concept that denotes a loosely tied family of properties. The cluster we propose is not meant as a definition of cognition, but as *diagnostic tool* to pick out systems that may suitably be viewed as cognitive[2].

P(S) contributes to S's cognitive abilities to the extent:

- *Condition AD (adaptability)*: P is involved in S's capacity to adapt its behaviour to match changing environments;
- *Condition IP (information processing)*: P is involved in S's processing information from its environment;
- *Condition I (intentionality)*: P makes structures in S about structures in its environment;
- *Condition C (consciousness)*: P makes S, or structures in S, conscious (in some of the various senses of the term).

These conditions are not disjoint, and they are not mutually exclusive. To insist, they are meant as diagnostic criteria such as one finds in, e.g., the Diagnostic and Statistical Manual (DSM) used by psychiatrists to diagnose mental illness. If a system has none of these properties, it will be said to be non-cognitive; it it has a few, it will be said to minimally cognitive; if it has all four, it will be said to be highly cognitive. Given both the notions of an emergent property and that of a cognitive property, we will define the notion of a *distributed cognitive property* as any system property that is both cognitive and emergent. Note that since both notions are matters of degree, and since both are independent, they form a space where various properties of systems may be plotted. We will refer to that space as the emergence/cognition space (E/C space). Temperature in a gas will fall squarely in quadrant

III of the E/C space, whereas the mass of the gas will fall in quadrant IV.



In the remainder of this position paper, we will explore various types of systems to see where they fall in the C/E space. Distributed cognitive systems, if there is such a thing, will be shown to fall in quadrant II.

3. A benign form of distributed cognition

The previous section was quite abstract ontological and conceptual stuff. The time has come to make things much more concrete by applying our definitions to specific cases. We start by a case that, we believe, is uncontroversially a case of distributed cognition (except that it is not the type of cases people usually have in mind when they talk about distributed cognition - we'll get to those in due time). The function of this section is to test our definition and set out a clear case of distributed (*qua* emergent) cognition before we address the more fuzzy cases.

Take the case of a formal neural network. It is made up of two types of (formal) entities: neurons (N) and connections (C) - we capitalize to mark the fact that we are talking of types of entities here. In typical standard models (such as multi layer perceptrons), neurons have a number of properties: an activation state (as), an integration function (if), and a transfer function(tf). To use the above formalism, N(as, if, th). They may have other properties, but let's stick to those to keep things simple. Connections, for their part, have a connection weight (w) and a connectivity <ni, nj>: C(w, <ni,nj>). Any specific neural network NNet of a large class of neural networks may be defined with only this simple set of entities and properties. Notice that none of NNet's component's properties deserve to be called "cognitive". Perhaps its integration and transfer functions can be called "computational" (a kind of analog computation), but this is not important here. Now, as everyone knows, systems such as NNet have been trained[3] to manifest various cognitive properties. Let's say NNet has been trained to categorize objects (for instance, faces). A qualitatively different type of property, CAT (for categorization) has emerged from the interaction of N's components. No hocus pocus here, as anyone can write the equation describing the emergent property of CAT, for an *n* neurons network:

- CAT(*N*) = *F*[*as*(n_i), if(n_i), tf(n_i); for i=1 to n; w_k, <n_i,n_j>(c_k); for (i, j) = (1,2) to (n-1, n) with i<j and k=1 to n(n-1)/2]

If, given its connectivity, NNet is a multi-layer perceptron[4], then *F* is simple linear algebra (vector by matrix multiplication, summation, etc.). Of course, psychologists who use neural networks to model and study categorization are not interested in this equation. They're interested in the global properties of the system. Does the net categorize this set of data, better than this one? Do its categorization ability generalize well? How is its ability influenced by time? To ask and answer all of these questions, they need vocabulary and measurements that apply to the global behaviour of the network. That is, they have a genuine epistemic project (understanding categorization), which can only be pursued at the level of the global system.

Where does S's categorization capacity fall on the C/E space? Before it was trained to manifest CAT, N behaved inadequately in its environment[5] (as measured by, e.g., by RMS error). CAT thus contributes to N being adapted to its environment (low RMS error). Depending on the neural network involved, learning may be ongoing throughout ontogeny, in which case, CAT will change over time to reflect the changing categorical structure or the environment, provided it does change. CAT also makes NNet sensitive to the categorical structure of faces in the environment. Given CAT, NNet can now behave in a manner that is sentive to that structure. Moreover, NNet has CAT because its hidden layer activation space has been partitioned in such a way that sub-spaces are correlated with particular faces in the environment. Whether this correlation is sufficient for full-blown intentionality, or whether it is only a minimal form of intentionality is a matter of debate. Nevertheless, CAT does satisfy condition I to some extent. Finally, CAT is not part of NNet 's (or NNet 's states) consciousness, since NNet is certainly not conscious. How about emergence? CAT is a typical emergent property. It fails condition IS: CAT is invariant when neurons are intersubstituted (neurons in ANN are usually clones) but certainly not when connections are. Note, and this is all very dependent on the exact nature of the ANN used, but in some case a few connections may be intersubstituted with only a slight decay in performance[6]. Hence CAT mostly (though not totally) fail condition IS with respect to its trained connections. CAT also fails

condition QS, RA and CI (for lack of space, we leave it to the reader to convince himself of this fact). It follows that CAT falls in Quadrant II, although it does not stand at its extreme (coordinate 1,1). Categorization in a neural network is a case of a *distributed cognitive property*. Of course, categorization in neural networks, as a case of distributed cognition, is unlikely to generate any debate. The difficult and problematic cases of distributed cognition are those where the system's components themselves have cognitive properties or are themeslves agents. But before we turn to full-blown agents, and to get a feel for the framework, let's take a case that is a little more challenging, that of swarm intelligence.

4. Distributed cognition between dummies: swarm intelligence (SI)

Distributed cognition in neural networks is uncontroversial. The case for distributed cognition becomes contentious when cognitive properties are said to emerge from the collective behaviour of agents. But the situation, we'll argue, is exactly parallel to the previous and thus should not raise any eyebrow. Let's start with simple agents, dummies really, before we move on to the really contreversial case of interacting agents holding cognitive properties.

Swarm intelligence (SI) is the name given to the global behaviour of large groups of locally interacting artificial agents or animals, such as bees building their hive, ants foraging for food, or birds and fish flocking. The components of SI systems are agents, which are endowed capacities (or properties). These agents and their properties, together with the way they interact given their properties, form the "agent level" of the SI system. The keyword in SI is *"simple"*. Agents are usually simple (simple rules, simple representations) and undifferentiated (though there may be a small number of categories of agents). As for most agents considered in the field of Multi-Agent Systems, their perception capacities are local: perception of their neighbour's behaviour or of traces left in the environment. Their action capacities are also local: moving, leaving traces in the environment, holding pieces of food. These local capacities completely determine their means of interaction with each other and with their environment. Yet, despite the purely local attributes of components at the agent level, global properties such as coordination emerge at the system level.

Consider a classical example of SI: ant foraging. All ants in an ant colony (AC) manifest three simple behaviours when they forage for food: (1) when no pheromone trail is present nearby, they move randomly; (2) when a pheromone trail is present nearby, they move toward it; (3) when a piece of food is present nearby, they pick it up and bring it back to the nest while leaving a pheromone trail. Now consider a single ant from the whole system engaged in foraging. The ant will move about randomly (behaviour 1) until it finds either a piece of food or a trail of pheromone. If it finds a piece of food, it will bring it back to the nest (behaviour 3) creating a pheromone trace between the food source and the nest. If, instead, it first finds a trail of pheromone, following it (behaviour 2) will lead it to a piece of food which it will bring back to the nest (behavour 3), *reinforcing the trace of pheromone that lead it to the food source*. As pheromones vanish from the environment after a while, only the trails that are positively reinforced by many ants will remain in the environment, that is, only the paths that link food sources to nest will remain. Morevover, if numerous sources of food are present in the vicinity of the nest, this phenomenon will particulary favour the closer food sources. From the collective behaviours of ants in the colony has emerged a global capacity: the capacity to *Find the path to the Closest Food source*, call it FCF(AC).

Is this global capacity of the ant colony a case of distributed cognition? FCF allows AC to adapt to its changing environement. As the first found food source is depleted, the next closest will be found and exploited, and so on. If a new food source "appears" in the environement (e.g. a cookie crumb is dropped by a human child), the same process will bring AC to preferentially exploit it if it's closest to the nest. Does FCF make AC sensitive to information in its environement, and does it allow AC to process that information? In some sense yes. FCF makes AC sensitive to its environement's food geography (or food distance metric). AC, given FCF, "knows" where the food sources are and what distance they stand from the nest. This nest-centric knowledge is only actual for the closest food source, but it is potential for all food sources. FCF is about the closest food source, in the minimal sense CAT above was about faces. Of course, like CAT, FCF is not part of AC's consciouness (for the same reason: AC is not conscious). Although we will not write the equation like we did above, it should be clear that this global capacity is emergent, just as the categorization capacity of neural networks is. FCF is invariant under intersubstitution of individual ants (which, like artificial neurons, are clones) and of phromones (both individually and positionally). Condition IS is thus satisfied. But condition QS is not as the number of ants in the colony cannot be reduced beyond a certain point (which is a function of the rate of depletion of the pheromone trace and speed of the ants). Conditions RA and CI fail also also to some extent. To explain the global behaviour of the ant colony, we need to use concepts such as "short path", "distance", but none of these are necessary to describe the behaviours of single ants. It follows from all of this that FCF is a distributed cognitive property that is just about as emergent as CAT but less cognitive.

Now let's take another example, where humans are being the dummies. Consider a campus with no paved footpaths. While the architects could have designed the layout of the paths through the lawn, it has been decided to wait for this layout to "come out" and pave it later. As students go from one building to another, they leave behind them not a trail of pheromone, but of crushed grass. If numerous students walk the same path, the grass begins to die, the path becomes apparent and attracts more and more students, with the same positive reinforcement phenomenon seen in the ant example (Brassac and Pesty, 1996).

Once again, let's ask ourselves if this example is a case of distributed cognition. Consider the systemic function *Design Footpath Layout* (DFL). DFL respects condition AD: if a cafeteria is closed or classrooms change place, DFL will display this new organisation of the environment. DFL mildly respects IP: information about the localization of buildings of interest to students and the connections between them is processed. It also respects I: DFL reflects the spatio-temporal structure of student's agenda. Finally, like FCF, it fails condition C. As for the position of DFL on the aggregativity-emergence of that function, it satisfies condition IS: like individual ants, students are interchangeable (with respect to DFL, of course!). DFL also satisfies conditions RA and CI to some extent. But fails condition QS: the dynamics of the system radically changes when the number of students involved falls below a certain threshold. So, from the standpoint of distributed cognition, DFL falls in the same category as the Find the path to the Closest Food source (FCF) property seen earlier. Note that the emergence/cognitive signature of the two systems is quite similar, which is why they are considered as two instances of the same type of systems, SI systems.

The rationale for the introduction of the second example is to show that it is very important to mark the difference between all the capacities that the agent has and those that they are actually using (Brassac and Pesty, 1996). While no one doubts that university students have

more cognitive capacities than ants, the same type of systemic functions can be obtained through two different implementations: one minimally cognitive agents in one situation (ants foraging) and one with full-blown human cognitive agents in the other (student pedestrians). The important point is that they actually play the same *role* in the system and thus in the emergence of the cognitive systemic property. When full-blown cognitive agents participate in SI systems, it is important to note the different cognitive properties manifested at the agent and the system level. When they participate in the SI system, students think about their next class, perceive their local environment (usually only a small portion of the campus ground), and locally decide where to make the next few steps. They are not thinking about designing footpaths or optimizing lawn to paved ground ratios. They're not aware of any of these system level SI properties. Nevertheless, just as in an ant colony, system level properties do emerge from their local behaviours, driven by their own individual desires and choices.

One might complain that such a process is resource consuming, compared to systems that manipulate explicit representations of the world (be they centrally controlled or decentralized) and where there are coordination processes and direct communication between agents. That is, one might think that explicitly increasing the agent's level of cognitive activity would be more efficient to achieve such goals (find the shortest route to a food source, or the best layout for the footpaths on a campus) than expecting them to be the emergent cognitive systemic property of a group of interacting unintelligent agents. There is no definite answer to this legitimate concern, as any answer would be strongly context dependant. An important point to note, however, is that simply increasing the cognitive capacities of the component agents does not guarantee an increase in the cognitive capacities of the system as a whole, since those are emergent and not aggregative[7]. Specifically, while agents with a high level of cognitive capacities spend time *computing* an optimal solution and *solving various* issues they wouldn't have if they were simpler (conflicts, synchronisation, etc.), dummy agents, such as ants, are already on the job. In some situations, cost of computing is higher than cost of acting[8]. Also, to be really complete on that matter of agent's cognitive properties and systemic cognitive properties, there are cognitive behaviours that are not emergent (mostly aggregative from its cognitive subparts) and there are emergent behaviours that are not cognitive (like trafic jams).

The important point in swarm intelligence is that global properties emerge without the need of explicit description of the global system at the agent's level. It is important to note that, as was the case previously with neural networks, the concepts needed to explain the behaviour of the SI system are not needed to explain behaviour at the agent level, even in the case of global properties that emerge from locally behaving humans (as in the lawn case). In the case of ants, the notion of a shorter path is beyond the analytical capacities of such agents in a reactive architecture. In swarm intelligence, like the ant colony example, agents usually coact: they do their local job from their own local agent's perspective. In multi-agent systems composed of higher-level agents, like a pack of wolfs, agents might collaborate: part of their (inter-)actions is not directly related to the task but to synchronisation or to resolution of conflicts. Somehow, collaboration needs some knowledge that there is a higher level task that is to be accomplished at a systemic level while this knowledge is not present for coaction, like it happens in swarm intelligence. This strength of SI systems is also its most serious drawback: how do you design agents and environments to get (or, as Varela would say, to enact) the desired emergent property? Two options can be considered: training the system (by adapting agent's models and organisation, as in artificial neural networks), or increasing complexity of their model to fit cognitive properties

in. Train them, and you loose the "no-differentiability" constraint which in turn leads to the loss of the robustness (as ants will have different role and status, taking one away is risky for the systemic property). Add more cognitive properties to the agents and not only the latter risk is still present, but also, as we pointed it out earlier, you end up with entirely new (emergent!) problems like coordination or conflict resolution. As such, SI systems are a transitional point between the "communities" of neurons that make up an artificial neural network and "communities" of cognitive agents enacting distributed cognition.

5. Distributed cognition between cognitive agents: multi-agent systems (MAS)

The previous examples (neural networks and swarm intelligence) show that cognitive systemic properties may emerge from the interaction of non-cognitive (neurons) or minimally cognitive (ants) components of systems. The footpath example, for its part, shows that, when the cognitive capacities contributed by the cognitive components of SI systems are separated from the rest of their cognitive capacities, humans may (unwittingly) contribute to the emergence of systemic cognitive properties. All these cases rate high on the emergence dimension but rather low on the cognitive dimension[9]. However, these systemic forms of cognition lay the groundwork for various system-level models of cognitive agents, such as Minsky's Society of Mind (Minsky, 1985), which follow a principle sometimes called "the recursion principle" in the MAS community: at a higher level of abstraction, a multi-agent system can itself be viewed as an agent. As we'll see, these models do not rate as high on the emergence dimension, taking sometimes a more aggregative stance, such as in a network of experts, each "stupidly" dedicated to a given task in a given context. The time has now come to turn to these systems. Our focus example will be the pack of wolves.

When the pack is hunting, each wolf behaves in way that gives the pack the behavioural capacity to surround its prey (SUR). No wolf in the pack has that capacity: each wolf can prevent the prey from going off in one direction, and "surrounding" is just a matter of locking every direction, which only the pack can do. With respect to emergence, SUR strongly fails two of the conditions, QS and CI (Qualitative invariance and Cooperation/Inhibition), as it is impossible to surround a prey with one wolf (QS failure) and as spatial organisation of the wolves greatly impacts this property. SUR is thus undoubtedly, while not completely, emergent. With respect to cognition, SUR validates to different extent three of the four criteria, failing only on the consciousness one (C): it clearly is adaptable[10] (condition AD) and it distributively generates actions for the system's subparts (wolves) according to a distributed perception (condition IP). The situation is less clear about condition I (Intentionality) but the exact extent to which this condition is respected will not impair the fact that SUR is a cognitive property of the pack. Surrounding a prey is thus an example of distributed cognition in a group of cognitive agents. The pack of wolves is an entity that has more fangs and claws than the lone wolf, but these are pure quantitative changes, giving it an aggregative physical property. But it also has the behavioural capacity to surround its prey, which is qualitatively new property, giving the pack an emergent cognitive property. However, although the wolves contribute all their cognitive capacities to the emergent behaviour of the pack, unlike human pedestrians designing footpaths, the wolf pack still shares with previous cases of distributed cognition the fact that it satisfies condition IS (intersubstituability): in hunting pack, wolfs with similar cognitive and behavioural capacities are interchangeable. What if this last aggregativity condition fails, that is what if agents are highly heterogenous, like a group of humans,

computers and various supporting devices? And what if it is humans who contribute their full cognitive capacities to the system's global behaviour? To end this position paper, we propose to address this last question.

Recall that, according to our definition of distributed cognition, true distributed cognition will occupy Quadrant II. Emergent cognitive properties are the sign of distributed cognition. Quadrant I, in which system level cognitive properties are mostly aggregative, is where, on this proposal, we find what Harnad (Harnad, 2005) calls collaborative cognition. In other words, distributed cognition is grounded on qualitative change, while collaborative cognition is grounded on quantitative ones. To get a feel for the distinction, imagine five persons decide to write a book together on a given subject S. The cognitive or intellectual project they set themselves to accomplish can be described as "Writing a book on S." In the extreme case, they might decide that the book is to be made up of five separate and distinct parts and decide that each is to write one part on her own, without any input from the other authors. The resulting book, they decide, will simply be the sum of part I through V. In this extreme case the resulting behaviour is almost entirely aggregative (almost as much as the weight of the pile of books). When it comes to collaborative cognition, this is an extreme case, almost to the point of non-collaboration. Of course, they still needed to collaborate to the extent of deciding how to divide up the work, which means that collaborative cognition is not purely aggregative (as Wimsatt point out, almost nothing in nature is). No one (one hopes!) collaborates in this fashion. Usually, each author will read, comment, re-write, etc., what the others have written. They will discuss points, debate and perhaps even argue on the book's specific content. Is that a case of distributed cognition? No, we may suppose, because for all the cooperation and inhibition going on (failure of condition CI), there is no failure of condition QS. If one author falls sick before completion of the project, the others can manage to complete the book. Or someone else may be brought in the project to replace the sick author (respect of condition IS). The book may not be as good a book (or it may be a better one!), but the authors may still manage *to write a book on S*. Nevertheless, all this collaboration/inhibition is bringing us closer to the fuzzy line between collaborative cognition and distributed cognition. Imagine now, perhaps *per impossible*, that each author has specific qualities such that no one else could replace her (failure of condition IS). If either one of them falls sick, then no book on S can be written (not now, not ever). The property “write a book on S” in this, perhaps impossible, case has crossed the fuzzy line between aggregativity and emergence. By our definition, we will have to qualify the collective behaviour of the authors as distributed cognition (Quadrant II).

6. Concluding remarks

A few points are worth mentioning here and we propose to address them in concluding. Note that the emergent property is very specific one: "writing a book on S". And in the extreme condition we presented above (each author has particular qualities that cause the failure of condition IS), that specific emergent property may be seen to be the result of distributed cognition. But can general cognitive properties (such as the capacity to categorize, perceive the environment, or recall something) that are qualitatively new emerge from human cognitive interaction? Before we address this question, let's explore the geography of Quadrants I and II. It can be noted that cognitive systems cluster in specific regions of the Quadrants. We argued that the cognitive properties of SI systems and current neural networks are quite emergent but do not score high as cognitive properties. They cluster to the far right a little above the abscissa (x-Axis). The cognitive properties of brains, we may

suppose, score high on both dimensions, with the cognitive properties of human brains scoring highest in the cognitive dimension. Finally, we saw in the previous section that the system-level cognitive properties of groups of humans cognitively interacting score high on the cognitive dimension but are mostly aggregative, with perhaps some very specific properties (e.g., writing books on S) crossing over the into emergence's territory. These properties thus cluster left of center, high above the abscissa. To reframe the previous question: Can human interaction generate cognitive properties that stand close to the (1, 1) coordinate?

We suspect these will be hard to come by. There are a few reasons for this. The first is simple. Humans already manifest many cognitive properties which are diverse and can be quite sophisticated (such as those required to pass the Turing Test). Accordingly, if, at the agent's level, humans make use of the full range of their cognitive properties, few system-level cognitive properties are likely to be seen are really new. Distribution as emergence sets the bar very high for what will count as distributed cognition in the human case: a qualitatively different cognitive property must emerge from the components' interactions. If no qualitatively different cognitive property can be found, then Ockham's parsimony principle urges us to err on the side of caution and refrain from adding new properties to our ontology where none is needed to account for the world[11]. However, remember that this is a fuzzy bar, because the notion of emergence (*qua* failure of aggregativity) is itself fuzzy. In cases that fall in the region where both concepts apply (like the 35 years old person which is a member (more or less) of the sets "young" and "old."), we believe *epistemic pragmatism* should rule, that is, counting the case as one of collaborative or distributed cognition rests in the final analysis on the pragmatics of explanation and understanding.

The second reason is linked to the concept of subjective emergence we introduced above. Imagine one truly qualitatively new cognitive property was to emerge through human interaction. Could we perceive and understand it? We might not perceive it simply because we are not looking at it[12]; or perhaps we are looking at it a wrong way; or again some of them might simply be imperceptible. These last cases are of no interest: if they have no direct or indirect perceptible effect, they might as well not exist. As for understanding these properties, as we pointed it out earlier, a qualitatively new cognitive property might simply be alien to us: it is difficult to imagine such a property, even if we are ready to put that label on properties we would otherwise consider magic, psychic, mystic or something like that (or even worse). Such a property we would perceive but not understand, at least at first. While speaking of swarm intelligence, we also made the following note: ant cognition is not able to use concepts related to the emerging property (short distance, etc.). We have to be opened to the possibility that there may be situations where human epistemic limitations prevent from understanding properties that emerge from our collective behaviour. In such a situation, even when perceived, the essence of the distributed cognitive property will remain out of our intellectual reach[13].

But the situation is not as bad as it seems. Even if *direct* perception of emergent cognitive properties might be problematic, indirect perception is much more likely. Dark matter cannot be seen, but its effects can be measured. Getting back to our wolf pack, the system-level analysis would be that *Surrounding* is more likely to prevent preys from escaping. At the agent-level (the wolf), the relevant disposition could be described as: "If, while hunting, I stay next to my fellow hunters but not to close, I get more food". Now, some superiorly intelligent wolf that would know of the notion of emergence might suspect that some emergent property is actually having an impact, while not being able to explicitly formulate

it. It remains a black box, but assumptions can be done with respect to it inputs and outputs. The same holds for collective cognition: some of the quantitative changes in systemic properties might come not only from aggregation of the same agent-level property, but also from emergent properties side-effects, and these can be perceived.

A last practical thought on distributed and collaborative cognition. In principle, a human with unlimited time and unlimited amount of pen and paper (or a very good memory!) could do everything a computer (or network of computers) can. But as soon as we stop theoretically arguing over the nature of various types of collective cognition, it is obvious that, since all of our capacities, including the cognitive ones, are bounded one way or another, the only option we got to increase cognition is with the help of sidekicks. Hence, the practical importance of studying interaction and dynamics of groups of humans cognitively interacting. This is especially true as the heterogeneity of the system's components increases (as in situations where humans and computers cognitively interact) from "simple" situations from the field of Human-Computer Interface (HCI) to that of Computer Supported Collective Work (CSCW), which the present authors used to collaborate, and all the way to what Licklider called man-computer symbiosis (Licklider 1960).

Notes

[1] We use the term here as in fuzzy first-order logic. Both distribution and cognition will be ascribed criteria to determine their membership function, that is, to what extent individuals (here individual properties and systems) falls in the concept's extension.
[2] As such, other candidate conditions have been discarded (like communication ability, autonomy, learning or expectations, behaviour based on goals, drives or intentions, qualia, etc.), and the actual choice is of course open to discussion.
[3] by algorithms that set an adequate value to property w of connections.
[4] In the case of a MLP, i, j and k shall be adapted in CAT's equation to show that the network is not fully interconnected but that neurons are layered.
[5] Of course, the notion of an environement in a neural network is very minimal. When the network is not embodied in a robot or animat, receiving environmental inputs through transducers and manifesing behaviours through effectors, its "environement" is reduced to input and output files, the structure of which rarely possess typical environmental structure like a topology.
[6] and sometime even with an improvement, as it might prevent overtraining and improve generalisation capacity of the network.
[7] It is worthy to remind here that the agents producing the layout of footpath have access to cognitive capacities, as they are humans.
[8] The swarm approach has other advantages, some related to the multi-agent approach, some related to the choice of simplicity at the agent level. The emergence of global properties from local perceptions and interactions have been shown to be efficent in situations in which the global system is simply unknown or too transient, such as packet routing in networks. The local nature of action and perception in SI systems, coupled with the absence of explicit or complex representation of the world, provides robustness: environment can be chaged, even dramatically, agents can be added and substacted (to some extent: remember this is an emergent property), and chances are that the global properties of the system will not be impaired.
[9] Brains, which we haven't discussed here, but which are presumably the ultimate neural network, would of course rate high on both dimensions.

[10] As a matter of fact, it can even be seen as a way of constraining the environment by leading the prey and the pack to a defined position, so that other functions can be used without the need for these other functions to be adaptable.
[11] Of course, Ockham would not have included *any* properties in his ontology, being a nominalist - nobody's perfect!
[12] For example, if we consider steel plates and the Size property, this property is part of every steel plate, but different organisation of the plates --i.e. Wimsatt's CI condition-- will change this property at the system level, making it a mildly emerging systemic property. But how would we recognise that some organisation, let's say, boat-shaped, would provide a qualitative emergent property, Floatability?
[13] Of course, if this inhability to understand is just a quantitative failure of our cognitive capacities, we can still rely on aggregation of our efforts (CC) to understand it.

References

Brassac, C. and Pesty, S. 1996. "La pelouse fourmilière; de la coaction à la coopération". In Quinqueton Muller, ed., proceedings of the *4èmes Journées Francophones sur l`Intelligence Artificielle Distribuée et Systèmes Multi-Agents* (JFIADSMA`96), pages 250-263, Paris: Hermès.

Baron-Cohen, S. 1997 *Mindblindness*. Cambridge: MIT Press.

Harnad, S. 2005. "Distributed Processes, Distributed Cognizers and Collaborative Cognition". *Pragmatics & Cognition*. http://eprints.ecs.soton.ac.uk/10997/01/distribcog.pdf

Hutchins, E. 1996. *Cognition in the Wild.* Cambridge: MIT Press

Licklider, J.C.R 1960. "Man-Computer Symbiosis" *IRE Transactions on Human Factors in Electronics*, volume HFE-1, pages 4–11, March 1960.

Minsky, M. 1985. *Society of Mind*. New-York: Simon and Shuster.

Nonaka, I and H. Tekeuchi 1995. The knowledge-Creating Company. Oxford: Oxford University Press.

Wimsatt, W.C. (1986). " Forms of Aggregativity." *In* Donagan, A.; Perovich Jr.; and Wedin, M., A. (1986). *Human Nature and Natural Knowledge: Essays Presented to Marjorie Grene on the Occasion of Her seventy-fifth Birthday.* Dordrecht: Reidel.