Reduction and emergence as features of cognition: unification and the extended mind hypothesis.

Raoul Gervais
raoul.Gervais@ugent.be

Merel Lefevere
merel.lefevere@ugent.be

Centre for Logic and Philosophy of Science
Ghent University
Blandijnberg 2
B-9000 Gent
Belgium

1. Introduction

Traditionally, reduction and emergence are thought of as features of reality. Roughly speaking, reduction in this ontological sense means that some entity, process or substance is nothing but an special aggregate or ordering of more fundamental entities, processes or substances; while emergence means that some entity, process or substance exhibits qualities or causal powers that go beyond those of the entities, processes or substances of which it is an aggregate or ordering.[i] To the extent that reduction and emergentism are mutually exclusive positions, the following dilemma occurs:

D – Either we accept reduction, which means giving up the belief in novel causal powers arising at higher levels of organization, or we accept emergentism, which means giving up the ambition to arrive at a unified, coherent body of knowledge.

To appreciate this choice as a dilemma, consider that neither horn is very attractive. First, we want our (folk) categories to denote entities and processes that have a meaningful impact in the world. This applies to our mental faculties and our emotions, to the arts, to intentions, to societies etc. Surely we want all these to make a difference. Accepting reductionism however, seems to rob them of their power to do just that. At the same time however, reduction resides at the very heart of

scientific progress. To reduce the complex to the simple, to see patterns in chaos, to bring together what was once separate etc., all this is central to increasing our understanding of the world. Yet accepting emergentism means accepting novel, unpredictable and unexplainable ('free floating') phenomena.

So, if we are reluctant to choose either horn, are there ways to avoid D altogether? Of course, there are other ways one can employ the concepts of reduction and emergence, and as we will see, applying them in different domains makes a difference to how D is formulated. In what follows, we will explore one such alternative interpretation of reduction and emergence, construing them as features of *cognition*, rather than reality itself. Specifically, we will argue for two claims. First, we will argue that:

i) It is possible to construe reduction and emergence as features of cognition (rather than features of the world or epistemic categories).

However, on its own this claim is rather unremarkable: if there is nothing to gain from construing reduction and emergence in this manner, i.e. if it solves no problem, then one might wonder why we should pursue this claim at all. Hence the second claim:

ii) Doing so allows us to dispel D *as a dilemma*.

Ad i), we will argue that it is possible to view reductionism and emergence as *cognitive values*, i.e. as reasons we have for construing and pursuing certain theories and explanations. Specifically, we will argue that both reduction and emergence can be understood as the tendency to pursue unification in our theories and explanations, although unification means something different in both contexts. Ad ii), we will argue that construing them so will allow us to adopt a pluralist position. On this pluralist account, reduction and emergence can be interpreted as two extremes on a continuum, the precise position on which depends on our interests at that moment. In this vain, D disappears as a dilemma, as we are no committed to one of its horn to the exclusion of the other.

Let us conclude this introduction with an overview of the paper. In section 2, we will provide some definitions and additional distinctions regarding reduction and emergence, so as to delineate these concepts more sharply. In section 3, we will briefly discuss their history as technical terms in philosophy, showing how their application has shifted from an ontological one (exemplified by the British emergentists) to an epistemic, or, as we will call it, theoretical one (in the twentieth century philosophy of mind). This gives us a first step to argue for i); a task we will set out to accomplish in section 4. Here, we construe reduction and emergence as features of cognition by interpreting them as cognitive values, i.e. as reasons to pursue those explanations that provide unification, although unification means something different in both cases. Consequently, we shall distinguish between what we call completeness-unification and simplicity-unification, or C- and S- type unification. In sections 5 and 6, we will consider both these types of unification in turn, showing how reduction and emergence, as cognitive values, lead us to seek out different explanations. In the concluding section, we will briefly recapitulate our progress and explain how our reinterpretation of reduction and emergence helps us to solve D, thus vindicating claim ii).

2. Some preliminary distinctions and definitions

Normally, reduction and emergence are viewed as contrary notions, to the extent that emergence is often understood as the absence or non-applicability of reduction and vice versa. For example, the entry of 'emergent properties' in the Stanford Encyclopedia of Philosophy has it that "…emergent entities (properties or substances) 'arise' out of more fundamental entities and yet are 'novel' or 'irreducible' with respect to them" (O'Connor & Wong 2009). Again, at a very fundamental level, this contraposition seems to impose a dilemma on us: accepting reductionism would entail sacrificing our mental properties as something special and distinctive, while opting for emergentism would entail giving up our cherished belief that, at bottom, everything is explainable in terms of fundamental physics.

Notwithstanding this supposed mutual exclusiveness of the two concepts, the picture just sketched tends to obscure the fact that historically speaking, the two terms have been applied to quite different things. Broadly speaking, one can distinguish between an ontological and an epistemological or theoretical reading of the terms. Let us first look at reduction.

- **Ontological**: *A is reducible to B iff all objects of the A-kind consist in nothing but objects of the B-kind*. This application of the term reduction is meant to capture statements like: fish tanks are nothing but collections of molecules. Put like this, reduction is a relation between substances. It is also possible to formulate a weaker ontological reductionism applying to *properties* rather than substances: *A is reducible to B iff all properties of the A-kind are nothing but properties of the B-kind.* To give a well-worn example, the property heat is nothing but mean kinetic energy. As substance reductionism is currently more an ideal than a widespread philosophical position, henceforth when we speak of ontological reductionism, we have in mind property reductionism, unless otherwise indicated.
- **Theoretical**: *theory A is reducible to theory B iff A is somehow shown to be a special instance of B*. Again, this can mean many things. It might be that A is deducible from B, or that from B we are able to predict A[ii], or A and B might even denote entire sciences rather than mere theories. For now, let us leave aside these technicalities. The main point is that the relata are no longer entities or properties, but theoretical descriptions of them. This sense of reduction is meant to capture statements like: claims about heat are nothing but special cases of claims about mean kinetic energy.

Conversely, emergence comes in two flavors as well:

- **Ontological**: *property A is emergent iff its occurrence is not determined by more fundamental properties*. This definition is meant to capture statements like: the behavior of the whole is more than the sum of its parts. For example: consciousness is more than just the collective activity of neurons.
- **Theoretical**: *theory A is emergent iff it is not explainable (predicted by, deducible from etc.) in terms of a more fundamental theory B.* For example: a biological theory referring to the

property 'being alive' is emergent if no amount of chemical or physiological theory can explain why some systems have the property of being alive.

By providing these definitions, we do not intend to delineate clearly the 'correct' use of the concepts of reduction and emergence, much less to explicate a set of necessary and jointly sufficient conditions for either of them to occur. Rather, we think that these preliminary characterizations of the concepts might help to get some grip on the subject before we embark on the task we set ourselves in the introduction. We do feel however, that they reveal certain features that have been historically associated with reduction and emergence. In the next section, we will (albeit, we admit, somewhat haphazardly) review some of this history to illustrate this point.

3. A brief history of the concepts: from ontology to philosophy of science

Arguably the first philosophic debate in which the concepts were used as technical terms was that between the mechanists and vitalists in the nineteenth and early twentieth century. Here, the concepts were used mainly in an ontological sense, with the mechanists claiming that life is at bottom determined by physical and chemical processes and laws, while vitalists like Hans Driesch maintained the existence of an entelechy operative in the organism; an element distinctive of living organisms and the organizing principle behind their movement, generation and procreation:

> …the actual organism, as it offers itself to observation, is certainly a combination of singularities, each of which may be described in terms of physics and chemistry, like a machine, and also all changes in these singularities lead to results which may be so described, but the reason of the *origin* of the combination and of all its changes is not a law or any combination of laws taught to us by physics and chemistry, but rests upon entelechy… (Driesch 1908, 137-138).

Between these two extremes, emergentists such as J. S. Mill and C. D. Broad held a middle position: while avoiding any reference to vital substances or entelechies, they nevertheless maintained that life could never be reduced to physical and chemical processes. For them, life was an ontologically emergent phenomenon, in the sense specified above. While there is ultimately one kind of stuff in the universe, i.e. matter, this matter can be arranged and ordered in numerous ways so as to allow for novel properties to emerge on higher levels of organization.[iii] Thus, reality is stratified into different layers, and the properties and entities residual to these layers, as well as the regularities governing them, have causal powers of their own, and are appropriately described in terms of a distinct scientific vocabulary.

This last point is important, for it shows how closely the ontological and epistemological interpretations of reduction and emergence are related. To be sure, emergentists like Mill and Broad entertained an ontological form of emergence, but their endorsement of this position affects more than reality, as it has consequences for the way we perceive, describe and engage with the world. To

put it bluntly, for the emergentists, ontological emergence entailed theoretical emergence, in that no amount of knowledge about the fundamental physical and chemical will account for higher-level properties:

> To whatever degree we might imagine our knowledge of the properties of the several ingredients of a living body to be extended and perfected, it is certain that no mere summing up of the separate actions of those elements will ever amount to the action of the living body itself (Mill 1856 p. 385).

Here, we see that the ontological stance has repercussions on the level of epistemology. In fact, the emergentist position on the specific issue of the theoretical (ir)reducibility of theories about life to those pertaining to physics and chemistry, generalizes to the more general query about the status of the special sciences in relation to fundamental physics, as C. D. Broad informs us:

> The question: Is chemical behavior ultimately different from dynamical behavior? seems just as reasonable as the question: Is vital behavior ultimately different from non-vital behavior? And we are much more likely to answer the latter question rightly if we see it in relation to similar questions which might be raised about other apparent differences of kind in the material realm (Broad 1925 p. 44).

Anyway, this shift of focus apparent in Broad's writing foreshadows the move in the philosophy of mind, occurring some fifty years after *Broad's Mind and its Place in Nature*, away from the specific question of the inter-theoretic relations between psychology and neurology, towards the general issue of the status of the special sciences as such. It is to this debate that we now briefly turn, as it is here that the concepts of reduction and emergence resurface.

In the wake of logical positivism with its eschewal of all things metaphysical, the philosophy of mind in the twentieth century saw a reformulation of the ontological mind/body problem into a problem concerning the relation between *theories* about the mental and *theories* about the brain. That is, specific stances one can take regarding the relation between mind and body, such as identity, functionalism (role/filler relations) or supervenience in one of its many guises, tended to be accompanied or even replaced by a stance on the relation between psychology and neuroscience, or more generally, between special sciences and fundamental physics. Thus, to give just a few examples, identity theorists turned to Nagel's model of theory reduction, functionalism was used to uphold the explanatory autonomy of the special sciences, while classic computationalists viewed the syntactic properties of formal systems operations, such as those realized by (universal) Turing machines, as undetermined by or emergent from their semantic properties. Although the more rigid positions such as Nagelian reductionism or Fodorian radical autonomy have been largely abandoned in favor of more tolerant, pluralistic stances, the debate between reductionism and anti-reductionism in the philosophy of mind is still very much alive, and the discussions of the preceding decades have resulted in a multitude of positions one can take on the issue of inter-theoretic relations.

All these positions and the various argument pro and contra need not concern us here. The point is that the debate about inter-theoretic relations can still be meaningfully interpreted in terms of reduction and emergence. Do the properties referred to by higher-level sciences have emergent causal powers (powers of their own) or are these powers simply inherited from their lower-level realizers? Are the generalizations higher-level sciences employ reducible to lower-level regularities or laws? Note also that the dilemma D, as sketched in the introduction, has remained, though in a different form: we still face the choice between excepting emergence, thus retaining the higher-level theories but thwarting our ambitions to arrive at a coherent body of knowledge, or pursuing these ambitions but relinquishing our belief in the explanatory success of higher-level theories.

With this new version of the dilemma explicated, we are now in a position to formulate more sharply what a reinterpretation of the concepts of reduction and emergence is meant to achieve: it must allow us to pursue the integration of knowledge into one coherent framework, while not jeopardizing the explanatory value of higher-level theories. Only then will we have succeeded in arguing for claim ii). However, before setting out to accomplish that particular task, we first need to argue for i), which is the object of the next section.

4. Reduction and cognition as features of cognition: two types of unification

In the previous section, we have illustrated that the meaning of the concepts reduction and emergence has shifted over time. The important point to remember for the present discussion, is that the turn from the old mind-body problem towards philosophy of science, in particular the issue of inter-theoretic relations, marks the transition from an ontological to a theoretical reading of reduction and emergence. From the perspective of the claims we made in the introduction, this transition is of signal importance, since it means that reduction and emergence are now applied, not only to reality as it is, but also to epistemic categories of our own making. Still, this is only half the story: the true question is whether these same concepts can be meaningfully applied to human cognition itself (claim i) and whether this application allows us to tackle the (now reformulated) dilemma D (claim ii). Here, we will pursue claim i). How then might reduction and emergence be interpreted as features of cognition?

Our proposal is that reduction and emergence might be understood as two *cognitive values* or *epistemic interests*, in the sense in which these terms are used by Longino, Lacey, Laudan, Douglas and others in the discussion on the question of whether science is value-free or not. Now in this discussion, cognitive values are contrasted with non-cognitive values, where the former are values that are concerned with the scientific methodology (predictive and explanatory power, simplicity, fruitfulness etc.) while the latter are all kinds of ethical, social and political values. Of course, the main issue when it comes to the supposedly value-free nature of science is whether these non-cognitive values should play a role in science, more specifically in the context of theory-acceptance, but this discussion is immaterial here: as we construe reduction and emergence explicitly as a *cognitive* value, our position is compatible with whatever answer one might prefer to this question.

We have in mind the following: reduction and emergence can be understood as the cognitive values to pursue those explanations or theories that provide *unification*. But how to understand unification?

In our view, the intuitive understanding we have of unification is actually ambiguous. For example, this is what Hempel has to say about unification:

> What scientific explanation, especially theoretical explanation, aims at is (…) an objective kind of insight that is achieved by a systematic unification, by exhibiting the phenomena as manifestations of common underlying structures and processes that conform to specific, testable basic principles (Hempel 1966 p. 83).

While according Kitcher, science:

> …advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same pattern of derivation over and over again, and in demonstrating this, it teaches us how to reduce the number of facts we have to accept as ultimate (Kitcher 1989, p. 423).

Now there are some important differences between these two characterizations of unification. For one, Hempel talks about phenomena in general while Kitcher applies the term to patterns of derivation. However, there are also two common themes clearly discernible. First, unification is aimed at bringing the many back to few, and second, this 'bringing back' provides some sort of *knowledge*, called 'insight' by Hempel, and 'understanding' by Kitcher. This understanding is constituted by a move away from the particular towards the general, or as Hempel puts it, the 'underlying structure'. In effect, it aims at providing the most *complete* description of the phenomena to be explained. Often, this will involve formulating more general laws or constructing a model. As completeness of description is what this type of unification is directed at, let us dub this type-C unification. Thus, reduction as a cognitive value is the tendency to look for those theories and explanation that provide type C-unification.

Yet this only partly covers the intuitive notion we have of unification. Sometimes, the bringing back the many to a few is not about explicating general laws, models or derivation patterns to capture the many, but about *reducing the complex to the simple*. Here, the aim is not to arrive at understanding by incorporating a lot of phenomena, but to chisel away those data as *irrelevant noise*. One treats as unified some phenomena, not because it is warranted by some underlying structure, but because it allows us to control and predict behavior that is otherwise meaningless. In this vain, we might for instance talk about a person has having beliefs and intentions, or of societies as exhibiting goals and desires. Unlike with C-type unification, these beliefs, intentions and desires etc. do not arise out of some underlying structure we know about (understanding or insight), but we attribute them simply to arrive at useful predictions and interventions. The details underlying these processes might be there, but we treat them as irrelevant noise and abstract away from them. Nevertheless, this attribution of intentions and desires does unify the phenomena to which they are attributed into conceptual collections or aggregates. As the target here is not completeness, but rather simplicity, so let us call this type S-unification. Thus, emergence as a cognitive value is the tendency to look for those theories and explanation that provide type S-unification.

Thus, our general argument for claim i) of the introduction is that reduction and emergence can be understood as features of human cognition if we construe them as cognitive values, that is, as the cognitive tendency to construct theories and explanations that provide C- and S-type unification respectively. As of yet, this proposal is of course highly abstract. In the next two sections therefore, we will try to flesh it out by providing a clear example for both reduction and emergence.

5. Reduction as cognitive value: psychology, neurology, and genetics

We have now construed reduction as a cognitive value which prompts us to construct explanations providing C-type unification. But how does this work in practice? In this section we will illustrate this with an example. Suppose, for example, that a correlation is found between a child's domestic situation and the probability of them developing clinic depression later in life, in the sense that in a given population P, it is found that those members coming from a broken home genuinely increases the chance of suffering from depression in adult life. About this correlation, one might ask the following question:

Q1: Why do members of P who come from broken homes have a higher chance of developing depression than those who do not?

Now this question might be answered either in traditional psychological vocabulary, e.g. by referring to emotional instability due to childhood trauma, or in neurological (that is, reductive) terms, e.g. by referring to serotonin imbalance destabilizing dopamine uptake in the prefrontal cortex. That is, the way this explanation-seeking question is formulated as such seems to have no bearing on the question whether we choose a reductive or non-reductive explanation. Of course, the neurological answer exhibits type-C unification, in that it gives us the most amount of insight into or understanding of the condition: it unifies the higher-level psychological or emotional conditions with underlying neural conditions, something the psychological explanation on its own does not accomplish. If in asking Q1 then, we are (primarily) motivated by the cognitive value of reduction, we had better choose the neurological explanation.

But of course, this is hardly a principled choice. There might be good reasons to prefer the psychological explanation. For example, in the context of intervention, it might be that we lack the means to intervene upon serotonin imbalance directly, while we have reasonable success treating emotional instability by applying psychotherapy etc. Suppose this situation applies, and that our initial explanatory question was driven the cognitive value of practical prevention, so that we choose the psychologically formulated explanation. This explanation might result in practical prevention policy: we apply preventative psychotherapeutic treatment to those members of P who are come from a broken home, that is, to members of population P*, which is a subset of P. Although this might be successful in preventing clinical depression to develop later in life, it is unlikely to be a hundred percent successful: some children from broken homes, will still develop depression, despite the treatment. In this case, a further explanation-seeking question arises:

Q2 Why do some members of P* develop depression despite having undergone preventative therapeutic treatment?

Furthermore, remember that although coming from a broken home is a probability raising factor for developing depression, this does not equate to a sufficient cause, so that at least some people who come from broken homes and have undergone no preventative therapeutic treatment do not develop depression. Thus, a third question arises:

Q3 Why do some members of P* not develop depression, despite not having undergone preventative therapeutic treatment?

Note that unlike Q1, Q2 and Q3 are questions that are explicitly formulated against the backdrop of the cognitive value reduction: answers providing C-type unification are the most appropriate in this context, while those referring to emotional states or environmental factors are far less so. The reason is simple, we want the answer to generalize to as much members of P* as possible. The appropriate answers then, are those who explain the differences observed between the subpopulations of P*, which will be the ones that unify the different occurrences and non-occurrences of depression as manifestations of some 'underlying structure', to borrow Hempel's words again. In the case under consideration, an appropriate explanation might refer to genetic differences among the members of P*: some people are genetically prone to developing depression and so might develop the condition despite the treatment, while others are genetically resilient to developing depression, so that they do not develop it even though they have not received treatment. We can now appreciate the difference between C- and S-type unification: though the genetic explanations clearly provide unification, they do not so by reducing the complex to the simple.

Anyway, this should suffice to give us an idea of how reduction as a cognitive value might lead us to construct certain explanations and not others, either directly, or by suggesting further explanation seeking-questions that are themselves most appropriately answered by those explanation that provide C-type explanations.

6. Emergence: the functional behavior of cells and the extended mind hypothesis

So far on reduction and C-type unification. How about emergence? Again, the claim is that emergence as a cognitive value leads us to pursue those theories and explanations that provide S-type unification, that is, explanations that simplify by attributing higher level qualities to certain aggregates or systems.

Let us look at an example from cell biology: the function of biomolecules in cells. A cell can be conceived as a species of macromolecule, i.e. as a specific type of aggregate of molecules. Now like other macromolecules, because of its chemical properties, a cell can react with an infinite set of

possible molecules. However, cells are different than other macromolecules in that although they could in principle react with an infinite set of possible molecules, as a matter of fact its molecular species only reacts with a few very specific molecules, making the set of actual reactions relatively small. About this observation, the following explanation-seeking question might arise:

Q4 Why do certain types of molecules in cells only exhibit a small number of possible reactions, while similar types of molecules in other macromolecular organizations have an infinite set of possible reactions?

In other words, what is so special about the cell? How does this specific environment constrain the number of possible reactions? Here, answers that refer to individual molecules, their chemical properties or even the atomic structure of the molecules themselves would not be appropriate responses. Looking at the 'underlying' structure is of no help, and consequently, C-type unification is not what is aimed at. That is, reduction, which leads to precisely those type of answers, seems not to be the cognitive value underlying Q4.

Instead, Q4 seems to be asked with emergence in mind. Emergence as a cognitive value prompts us to seek out those answers that are unifying in that they attribute systemic qualities or properties to the phenomena in question, and abstract away from the details (noise) which would be cited by a reductive explanation. In the case under considerations, the odd behavior of molecule-types in cells can be explained by ascribing functions to the molecules. The number of possible reactions is limited because it is only those reactions, and not others, that define the functions these molecules have within the overall structure in the cell. For example, having a metabolic function, in relation to the self-maintenance function of the overall cell, will restrict the number of possible reactions to congruent to a specific metabolic process.

The point is that this explanation makes sense of the behavior of molecule types within cells (as opposed to other macromolecules) by attributing functions to these molecules, and indeed the cell as a whole. It interprets behavior as aimed at fulfilling some goal, even though from a 'deeper' level, these interpretations and attributions are quite unjustified. Thus, the explanation provides type-S unification: it groups together and treats as unified a specific range of molecules and chemical reactions, and by doing so simplifies what appears to be a very complex process. Thus, as Q4 is asked with emergence as the primary cognitive value, one can see why this particular answer is appropriate: because it provides type S-understanding.

Of course, this type of explanation is not limited to cell biology. Type-S understanding is sought after in many scientific disciplines. It allows us to understand, predict and even manipulate, in simplified terms, the behavior of complex systems such as swarms, intestines and societies. In this vain, Andy Clark's *extended mind hypothesis* points to perhaps the most radical way in which cognition might be understood as emergent: although cognitive capacities might typically arise in heaps of grey cells that have been grouped together in a certain way, they might also arise in the interplay of brains and artificial ingredients (Clark & Chalmers 1998). To use Clark's example, if a patient suffering from Alzheimer consults a notebook to bypass having to draw on his impaired memory, then this notebook forms a proper part of the cognitive activity of memory-retrieval. The important thing is that we develop ontological flexibility, conceptually grouping together entities and activities as systems that we normally would not think of in these terms. This ontological flexibility is

traded in for explanatory simplicity, which is just what emergence, understood as a cognitive value, aims to achieve.

Conclusion: pluralism and the continuum of interests

To briefly recapitulate, we have seen that in the twentieth century, reduction and emergence began to be chiefly applied, not to reality, but rather to the epistemic categories we use to describe this world. As we have said, this is already an important step toward claim i): the idea that reduction and emergence can be meaningfully applied to human cognition itself. We have argued for this claim by reconstructing them as cognitive values, each of them aimed at a different type of understanding. By considering some examples for each, we have hopefully given the reader some idea of how this would work with respect to actual scientific explanations or theories. However, the mere possibility of construing reduction and emergence in this manner is not exciting by itself. Only if construing them so would confer some added benefit on us, can such an exercise be deemed philosophically worthwhile. To remain within the terminology of this paper, we have to tackle dilemma D in order to argue for claim ii). Then, have we made any ground so far as this second challenge is concerned? Let us make some brief remarks on this issue at the close of this paper.

What the above examples have shown us, we think, is that what type of explanation is appropriate depends on the context. This position comes down to explanatory pluralism, which, as we understand it, is the claim that judgments about the appropriateness of an explanation or theory ought to be made, not on the grounds of a priori theorizing, but by taking into account the cognitive values that the explanation in question is meant to serve, and consequently varies from case to case. If our cognitive value is reduction, as in the genetic explanation explored in section five, we are most likely better off with an explanation that provides type-C unification, while if it is emergence, an explanation yielding type-S unification is to be preferred. If reduction is the only cognitive value in play, we might want to lay bare the underlying structure of a system, by describing the laws or generalities under which it operates, or by decomposing it and constructing a model of it. If we are solely after emergence, then the implementational details are just noise to us and we want to stick with attributing certain system-level properties to the aggregate under consideration. However, these two types of unification are not mutually exclusive and can even complement each other: we might want some implementational details to accompany the emergent system we are describing. Thus construed, C- and S-type constitute the extreme ends of a continuum, with the cognitive values of reduction and emergence respectively pulling towards those ends. A given explanation will often represent a tradeoff between these two pulling factors. Thus, the place on the continuum we will ultimately end up is where reduction and emergence balance out for the explanatory task at hand.

But then, barring the two extreme positions, most explanations will often be driven by both cognitive values. On this account, dilemma D as we sketched it in the introduction disappears as a dilemma: we can both pursue the ideal of arriving at a coherent body of knowledge, while at the same time accepting that the level of nitty-gritty details allowed may not be exhaustive. As philosophers of science, the only interesting question left is which of these ideals gets more weight in a given situation and why. This constitutes our case for claim ii) of the introduction.

References

- Broad, C. D. (1925). *The Mind and its Place in Nature* (1th edition). London: Routledge & Kegan Paul.
- Chalmers, D. J. (2006). Strong and weak emergence. In P. Davies & P. Clayton (Eds.) *The Re-Emergence of Emergence*. Oxford University Press, pp. 244-256.
- Clark, A. & Chalmers, D. J. (1998). The extended mind. *Analysis 58*: 10-23.
- Driesch, H. (1908). *The Science and Philosophy of the Organism Vol. II.* London: Adam and Charles Black.
- Hempel, C. (1966). *Philosophy of Natural Science*. Englewood Cliffs: Prentice-Hall.
- Kitcher, P. (1989). Explanatory Unification and the Causal Structure of the World, in *Scientific Explanation*, P. Kitcher and W. Salmon, 410-505. Minneapolis: University of Minnesota Press.
- Mill, J. S. (1856). *A system of Logic* (4th edition). London: John W. Parker and Son.
- O'Connor, T. & Wong, H. Y. (Oct. 23, 2006) "Emergent Properties", *The Stanford Encyclopedia of Philosophy (Spring 2009 Edition)*, Edward N. Zalta (ed.), retrieved 20 October 2011 from: URL = <http://plato.stanford.edu/archives/spr2009/entries/properties-emergent/>.

---

[i] In section 2 we will introduce some more precise distinctions and definitions for these concepts.

[ii] Regarding reduction as deducibility or predictability, David Chalmers (2006) distinguishes between strong and weak emergence to capture the impossibility of these two types of reduction respectively. He argues that it is important to keep these two types of emergence distinct, as strong emergence entails weak emergence, but not vice versa. In this paper however, we are interested in the difference between ontological and epistemological (or what we term theoretical) interpretations of the notions of reduction and emergence, and on this distinction, both deducibility and predictability clearly reside in the sphere of the latter.

[iii] Thus, late 19th- and early 20th century emergentism can be seen as the precursor to the contemporary and now dominant position of non-reductive physicalism in philosophy of mind.