
Levels

Carl F. Craver

The levels metaphor is commonly used to describe science, its theories, and the world. Yet the metaphor means different things in different contexts, inviting equivocation. These distinct applications of the metaphor can be distinguished by the relata they relate, the relation between levels that they assert, and the rule by which they locate items at a level. I discuss these many applications of the levels metaphor with an eye to developing a descriptively adequate account of one particular application: levels of mechanisms. I argue that this application of the metaphor is central to the explanatory practices of the special sciences and defensible as a metaphysical picture of how phenomena studied in the special sciences are constituted.

Keywords

Emergence | Explanation | Hierarchy | Interlevel causation | Mereology | Reduction

Author

[Carl F. Craver](#)

ccraver@artsci.wustl.edu

Washington University
St. Louis, MO, U.S.A.

Commentator

[Denis C. Martin](#)

denis.martin@hu-berlin.de

Humboldt Universität zu Berlin
Berlin, Germany

Editors

[Thomas Metzinger](#)

metzinger@uni-mainz.de

Johannes Gutenberg-Universität
Mainz, Germany

[Jennifer M. Windt](#)

jennifer.windt@monash.edu

Monash University
Melbourne, Australia

1 Introduction

The levels metaphor is ubiquitous in our descriptions of science and the world. So simple and elegant, the metaphor takes an apparently heterogeneous collection of objects and arranges them in space from bottom to top. The metaphor works in so many contexts because it leaves open just what kinds of object are to be arranged, what distinguishes top from bottom, and what it means to say that an object is at some levels and not others. This flexibility explains the metaphor's fecundity, but it also helps to obscure the fact that it is used in many different ways in many different contexts.

A survey of kinds of levels drawn from science and philosophy would have to include levels of abstraction ([Floridi 2008](#)), aggregation ([Wimsatt 1997](#)), analysis ([Shepherd 1994](#);

[Churchland & Sejnowski 1992](#)), causation and explanation ([Kim 1998](#)), implementation ([Marr 1982](#)), organization ([Churchland & Sejnowski 1992](#)), processing ([Craik & Lockhart 1972](#)), realization ([Gillett 2002](#)), sizes ([Wimsatt 1976](#)), sciences, theories, and explanations ([Oppenheim & Putnam 1958](#)). Many of these familiar applications of the levels metaphor are distinct from but also clearly related to one another. And when they are related, they often have rather indirect and reticulate connections. The level metaphors thus takes subtly different forms when applied in neighboring contexts, and this obscures the extent to which features of one application of the metaphor do and do not transfer from one context to the next. My first thesis, then, is that our ways of describing sci-

ence and the world contain many distinct, legitimate applications of the levels metaphor that are either unrelated or that have only indirect relations with one another.¹

This *descriptive pluralism* about the levels metaphor is directly opposed to eliminativism about levels (Fehr 2004; Machamer & Sullivan 2001; Thalos 2013). The suggestion that we might be better off abandoning the levels metaphor is about as likely to win converts as the suggestion that we should abandon metaphors involving weight or spatial inclusion. These metaphors are too basic to how we organize the world to seriously recommend that they could or should be stricken from thought and expression. Yet, descriptive pluralism about the levels metaphor is consistent with the thought that some applications of the metaphor distort the structure of the world or represent it as having conceptually incoherent structures. I discuss some examples below. The central message of this paper is that there can be no single verdict concerning the utility or conceptual soundness of the levels metaphor *simpliciter*. The metaphor must be evaluated and used with caution, especially when it is called on to settle disputes about the character of science and the metaphysical structure of the world.

As some motivation for adopting this proposal, and as a step toward a more positive thesis, I show that we can avoid some simple confusion by separating the different applications of the metaphor. To make this case, I build slowly toward a particular application of the metaphor that, as I have argued elsewhere (Craver 2001, 2007), is central to explanatory practices in neuroscience and across the special sciences: levels of mechanisms. This application of the levels metaphor is metaphysically plausible and, so far as I can tell, more or less innocuous; that is part of its virtue. Yet this simple and useful application of the metaphor can begin to appear problematic when it is inappropriately assimilated to other applications that

serve altogether different purposes in our thinking about science and the world.

My point is not to defend levels of mechanisms as the one true application of the levels metaphor (that would be as pointless as eliminativism). Rather, my first positive goal is to provide a reasonably clear account of levels of mechanisms and to show that this application is metaphysically benign yet exceptionally important for doing science. Levels of mechanisms are, as would be expected, richly but indirectly connected with many other applications of the metaphor. My second goal is to highlight and disentangle some of the confusions that arise from failing to keep levels of mechanisms distinct from other senses of levels. In particular, I show that commitment to the existence of levels of mechanisms entails no commitment to: a) monolithic levels in nature, b) the stratification of sciences by levels, or c) a tidy hierarchy of theories among the sciences. I will also show why levels of mechanisms are d) distinct from Marr's views about levels of abstraction and e) distinct from levels of realization more generally. I argue that f) the idea of interlevel causation is conceptually awkward within levels of mechanisms (but not to levels of size, for example). Furthermore, g) the idea of levels of mechanisms nicely expresses the idea of emergence as a kind of non-aggregativity while providing no support to those who seek evidence in biology for a more robust kind of emergence. The failure to disambiguate altogether separate applications of the levels metaphor creates a conceptual malaise for which levels of mechanisms are at least a partial cure.

2 Refining the levels metaphor: Three defining questions

In its barest of forms, the levels metaphor demands little of its object; it requires only a set of items and some criterion for ranking them as higher or lower than one another in some respect. Seniors are at a higher level in the American high school system than juniors, poetry is at a higher level than pushpin, lust is at a higher (and lower) level than like, and cells are at a higher level than molecules. In these ex-

¹ Standard etymologies trace the term "level" to the balance and from there to the idea of a flat, horizontal landing, as in the stories of a building. From there, it is easy to see how the metaphor might be extended to the kinds of hierarchy discussed in this paper.

amples, it is obvious that different kinds of thing are related by entirely distinct kinds of relation. In subtler cases, the equivocation is less noticeable, and for that, all the more misleading.

Three defining questions can be used to explicate how the levels metaphor is applied in a given context:

The Relata Question: What kinds of item are being sorted into levels?

The Relations Question: In virtue of what are two items at different levels?

The Placement Question: In virtue of what are two items at the same level?

The Relata Question provides an important clue about the intended sense of levels. The flexibility of the metaphor allows it to be applied to *abstracta*, such as branches of mathematics and ethical principles, or to *concreta*, such as astronomical objects and stereo equipment. The metaphor can be applied to types, such as sergeants and corporals, or to tokens, such as the relationship between Colonel Blake and Corporal O'Reilly. It can be applied to objects such as cats and mountains, to activities such as releasing neurotransmitters and making decisions, and to properties such as excitability or charge. Within the neurosciences, the levels metaphor is applied fluidly to causes, descriptions, developmental stages, events, explanations, scientific fields, objects, properties, techniques, and theories. Confusion arises when we assume that each application is the same.

The Relations Question concerns the ordering relationship by which items are said to be at a higher or a lower level than one another. A theory, for example, might be said to be at a higher level than a second if the first is derivable from the second (and not vice versa); the lowest-level theories are in this sense “fundamental.” Poetry might be said to be higher than pushpin in the sense that it requires greater intellectual skill and training to take pleasure in the former than to take pleasure in the latter. A technique might be said to be at a lower level than another because it detects phenomena at a smaller size scale. Some applications of the levels metaphor are discrete in the sense that there is a gap between things at lower and high

levels; other applications are continuous, as when one uses the metaphor to describe size. We are unlikely to confuse such wildly different kinds of relationship. However, as we will see, the metaphor is used in other contexts where it is beguilingly difficult to keep them distinct, even for those who know better.

The Placement Question asks for the principle by which different items are located on the same level. Many uses of the levels metaphor rely at heart on an answer to the placement question. When the metaphor is used to describe size scales, for example, puffins and porcupines are at roughly the same level, vasopressin and oxytocin are at roughly the same level, and hydrogen and oxygen atoms are together at a lower level still. Juniors are all juniors because they are in their third year of American high school. For Marr, computational level questions are directed at what is computed and why it is computed that way (Shagrir 2010; Bechtel & Shagrir 2013).² Not every account of levels requires an answer to the placement question affirmatively. Indeed, it is of central importance that the idea of levels of mechanisms articulated here entails no positive story about what it means to be at a level, only a negative story about when things are not at different levels.

3 From gesture to prototype

Perhaps the most common application of the levels metaphor is to gesture loosely at the relationship between different fields of scientific research, *levels of science*.³ In neuroscience, for example, some researchers work at “the molecular level,” doing things such as sequencing channel proteins, studying enzyme kinetics, or manipulating genes. Others work at the cellular level, doing things such as staining cells, recording action potentials, or studying neural migration. Others study brain regions, characterizing

² One consequence of the following discussion is that not every account of levels must offer a unique answer to the placement question. Levels of mechanisms are defined by their distinctive relata and relations; these constraints, by themselves, offer no unique answer to the placement question. This is why levels of mechanisms are, as I will argue, local rather than monolithic.

³ For the relevant sense of a scientific “field”, see Darden & Maul (1977) and Darden (1992).

their anatomical features or, studying the propagation of neural signals within them. Still others work at the level of systems, using functional magnetic resonance imaging (fMRI), transcranial magnetic stimulation (TMS), and cognitive tasks to find large-scale cognitive systems in the mind-brain. One could perhaps insinuate other levels between these, and one could certainly extend the hierarchy further down or higher up. But the central idea is that the *scientific fields* can be ordered as higher or lower than one another.

Scientific fields are individuated in part by their theories (Darden 1992). The gestural sense of levels, then, can seem to carry the implication that scientific *theories* are or will someday be ordered more or less clearly into levels. Oppenheim & Putnam's (1958) influential view of the unity of science is based on a rough correspondence between levels of science, levels of theory, and levels of parts and wholes (see Table 1). They divide the world into six ontological strata (societies, organisms, cells, molecules, atoms, and elementary particles). These strata are defined by mereological relationships among types: elementary particles are the parts of atoms, atoms are the parts of molecules, molecules are parts of cells, and so on. Each of these strata is assigned a distinct science: economics and the social sciences at the top, particle physics at the bottom. Each science develops its theory more or less autonomously from the others, so the theories developed by these sciences can themselves be ordered, like the layers of a cake, from top to bottom. The unity of science, for Oppenheim and Putnam, is to be achieved by explaining phenomena in the domain of a higher-level science, as described in the theory of that science, in terms of the items in the domain of the more fundamental science, as described in the theories of that science. (Levels of mechanisms, as defined below, involve a kind of part-whole relation as well but without any commitment to the idea that such type-level part-whole relationships correspond in even a rough way to the structure of the sciences or to the structures of their theories.)

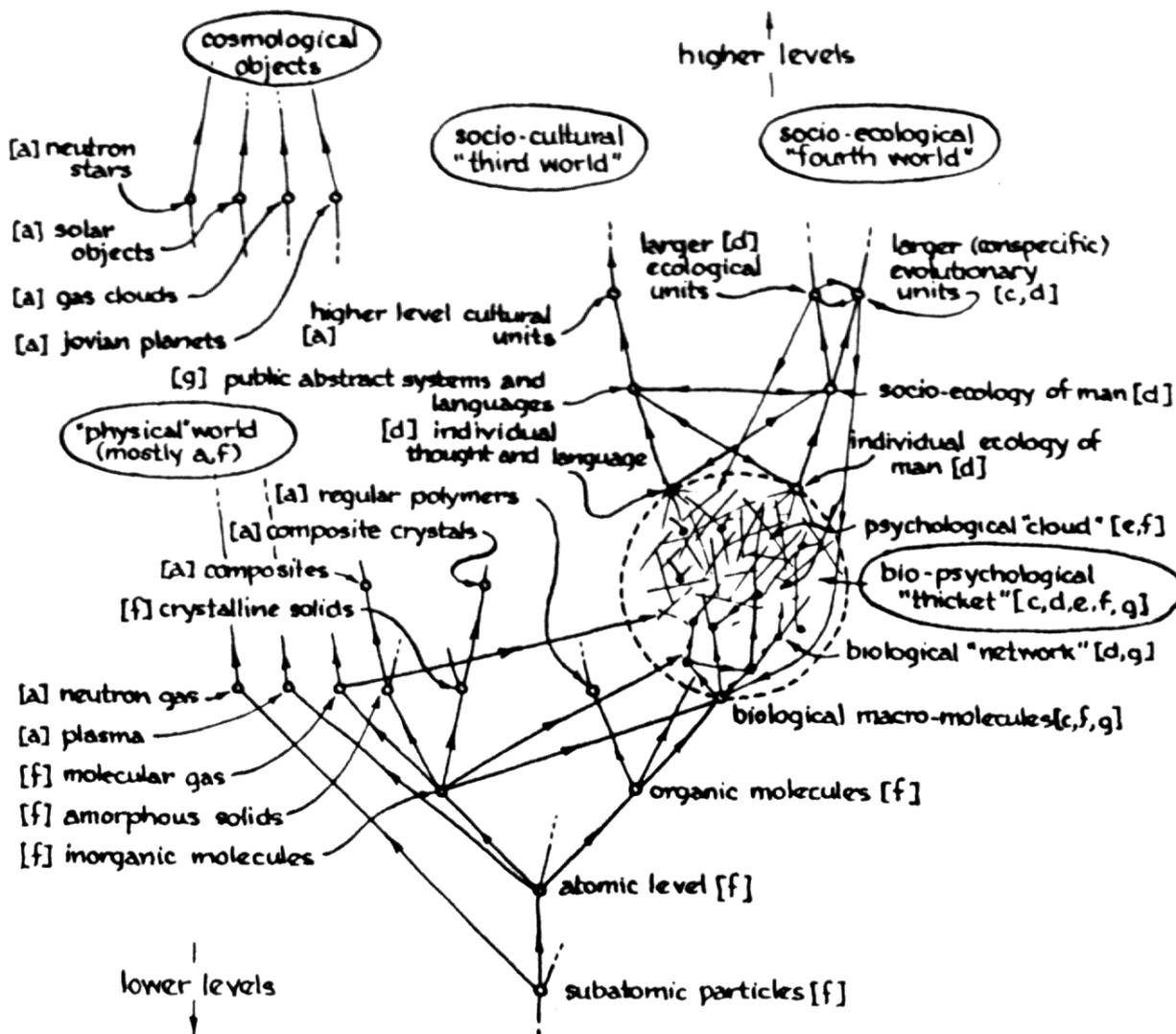
Wimsatt's detailed and influential exploration of the levels metaphor confronts Oppen-

heim and Putnam with the complexity of the levels found in many areas of contemporary science (Wimsatt 1976). Against Oppenheim and Putnam's six-layer model, Wimsatt's "Reductionist Illustrative" (Figure 1) represents multiple branches of levels fanning out from the lowest level in subatomic particles to cosmological objects, the sociocultural world (e.g., economic and political phenomena), and the socioecological world (e.g., evolution).

Wimsatt's tree diagram, however, represents only one aspect of his *prototype account* of levels that encompasses many more features than Oppenheim and Putnam's layer-cake mapping in Table 1. The core features in Wimsatt's prototype are:

- **Size.** Higher-level items are larger than lower-level items.
- **Composition.** Higher-level items are made up of lower-level objects and processes.
- **Laws.** Laws of nature hold only or mostly between items at the same level.
- **Forces.** Distinct forces operate at different levels.
- **Predictability.** Levels are local maxima of regularity and predictability that appear at different size scales.
- **Detection.** Items at a given level tend to be detected or detectable primarily by other items at that level.
- **Causes.** Causal relationships hold only or mostly between items at the same level.
- **Theories.** Scientific theories describe phenomena exclusively or mostly at a single level.
- **Techniques.** Different techniques and instruments detect items at different levels.
- **Disciplines.** Different disciplines of science direct their attention at different levels.

Wimsatt's view is a prototype view in the sense that it characterizes the levels metaphor in terms of a core set of features, not all of which must be present in order for the metaphor to apply. Insofar as Wimsatt embraces a prototype model, he can be seen as embracing descriptive pluralism while, at the same time, holding that there is a sufficiently strong family resemblance



(h) A reductionistic (?) illustrative* phylogenetic ontology of our world as we see it.
 [letters in parentheses refer to local character of network around that node]

* A diagram like this is obviously highly tentative. Although it has been constructed with many specific relations in mind, I would claim accuracy only of a general qualitative sort - i.e., for the rough general distribution of network properties, as indicated, e.g., by the distribution of letters. W.C. Wimsatt 1973

Figure 1: Wimsatt's tree of levels branches to preserve compositional relationships among levels (1976).

among the plurality of applications of the levels metaphor to warrant their inclusion in a single prototype.

Is the levels metaphor sufficiently unified across these different applications to warrant a single prototype? My remarks on the relation, and placement questions should already indicate that it is not—that different features in Wimsatt's list are at best indirectly related and

so fail to map to one another in any tidy way. While the prototype approach usefully highlights the complexity of the levels metaphor, it also obscures the extent to which the different features in the prototype are features of different applications of that metaphor.⁴

⁴ Wimsatt's diagram in Figure 1 reflects this. The branching tree structure is ordered by compositional relations. Wimsatt's view of levels as dissipating waves (see Figure 3 below) flouts that relation.

Table 1: Oppenheim and Putnam's layer-cake sketch of the levels of mereology (left), sciences (middle), and theories (right).

Mereological level	Sciences	Theories
Societies	Economics	Classical Economics
Organisms	Psychology	Law of Effect
Cells	Cytology	The Neuron Doctrine
Molecules	Chemistry	The Central Dogma
Atoms	Physics	The Bohr Model
Sub-Atomic Particles	Quantum Mechanics	Schrödinger Equation

4 Levels of sciences and theories

Wimsatt, Oppenheim, and Putnam all include within their analysis of levels the gestural idea that different fields or disciplines of science are arranged by the sizes of the objects they study. Wimsatt's branching hierarchies depict a more ornate structure. Within that structure, it seems inappropriate to say that astrophysics is at a higher level than biology or economics, though astrophysicists typically deal with things that are orders of magnitude larger than the things biologists and economists study. The gestural sense of levels doesn't seem to branch that way.

When we apply the levels metaphor to *sciences*, the relata are units of scientific organization (such as fields, research programs, or disciplines). Answers to the relations and placement questions are more difficult to discern and are likely impossible to express both accurately and concisely for this application. Size seems to be relevant, but we have just seen that it cannot be the whole story. The branches in Wimsatt's diagram follow, in addition to size relationships, relationships of composition. The things studied by economists (groups) are composed of things studied

by psychologists (organisms), which are composed of things studied by physiologists (physiological systems), and so on. The things studied by Darwin are composed of the things studied by zoologists, which are composed of the things studied by cytologists. The point of these examples is not to get the branches in Wimsatt's hierarchy exactly right; any proposed hierarchy of the sciences and the items in their domains is bound to be historically contingent and provisional at best.

In fact, many sciences appear to resist tidy compartmentalization within levels. Neuroscience, especially cognitive neuroscience, is a paradigm of multilevel science, encompassing the study of ions, ion channels, cells, populations of cells, brain regions, and behaviors of whole organisms. No competent evolutionary biologist can avoid knowing something about genes, physiological systems, organisms, populations, and environments. Many sciences, in short, contain items within their domain that stand in compositional relations to one another. Such sciences often construct multilevel theories that integrate findings across multiple levels of organization. This is one reason why the relationship between levels of science and part-whole levels is indirect.

Another reason is that, in many cases, more than one science is dedicated to studying items at the same mereological or size level. Cytologists, anatomists, and electrophysiologists all study aspects of cells with different tools. The ethologist and the experimental psychologist study animal behavior, but they approach that behavior with different assumptions, methods, and theories. Economists, ecologists, epidemiologists, and organizational psychologists study populations of organisms. The relationship between levels of science and the ontological levels that Oppenheim and Putnam presume is many to many.

For this reason, it is unlikely that any precise answer to the placement question will correctly express the application of the levels metaphor to sciences. One might say that two sciences are on the same level when they pertain to items at the same compositional level. Perhaps it makes sense to say that Camillo Golgi was investigating the same level when he stained Purkinje cells with silver nitrate that Alan Hodgkin was investigating when he used his voltage clamp to study the action potential of the squid giant axon. They were both studying cells, but they studied different phenomena and used different techniques. If we focus now on the parts of these wholes, we see that these different scientists are not even on the same branches of a Wimsatt diagram, and the levels metaphor begins to break down. Ask Golgi about the relevant parts of the cell, and he will tell you about its gross morphological features and its organs. Ask Hodgkin and Huxley about the relevant parts of the squid giant axon, and they will tell you about membranes, axon hillocks, ionic conductances, and voltage gradients. An epidemiologist might talk about nodes and networks and hubs in a model of contagion. Economists will talk about producers and consumers. Differences in scientific interests often entail differences in the relevant ontology for the science; and the same thing can be carved into parts in many ways depending on what one is interested in describing or explaining (Kauffman 1971; Wimsatt 1972).

The take-home lesson: the application of the levels metaphor to fields of science yields a

notion of levels only indirectly related to ontological levels (as understood in a roughly compositional, part-whole sense). The idealized, Oppenheim-Putnam correspondence between levels of science, levels of theory, and levels of mereology breaks down in the face of this many-many mapping. And the compositional aspect in Wimsatt's prototype appears to be only loosely related to the application of the levels metaphor to fields of science. These are, in short, distinct applications of the metaphor, offering different answers to the relata, relations, and placement questions. As a result, an understanding of how sciences can be organized loosely into levels provides no direct insight into ontological levels. This will come as no surprise to those who study intellectual history, or to those who have witnessed for themselves how fields of science change their boundaries over time. Our age, perhaps more than any other, has witnessed an explosion of hybrid fields (neuroeconomics, behavioral genetics, cognitive ethology) that cross levels, combine approaches, and attempt to feed off insights shared between distant scientific neighbors. The historical relativity of disciplinary boundaries makes them unreliable guides to ontology.

The same considerations suggest that *levels of scientific theory* will also have a many-many relationship with ontological levels. In this application of the metaphor, the relata are theories or models. And the relationship is typically construed as a kind of subsumption, e.g., deductive subsumption (Hempel 1965; Schaffner 1993; Kitcher 1989), or some kind of similarity or inclusion (Bickle 1998). The disciplinary hodgepodge of the special sciences fails to match this philosophical reconstruction. Single theories, such as the theory explaining spatial memory in terms of memory systems, grid cell organization, synaptic plasticity, and changes in ionic conductances through a membrane (see Moser 2008), often reach across many different part-whole levels (Darden & Maul 1977; Bechtel 1988; Craver 2002, 2008). One and the same mereological unit (e.g., cells) can appear in many distinct theories (e.g., neurons play some role in most theories in neuroscience).

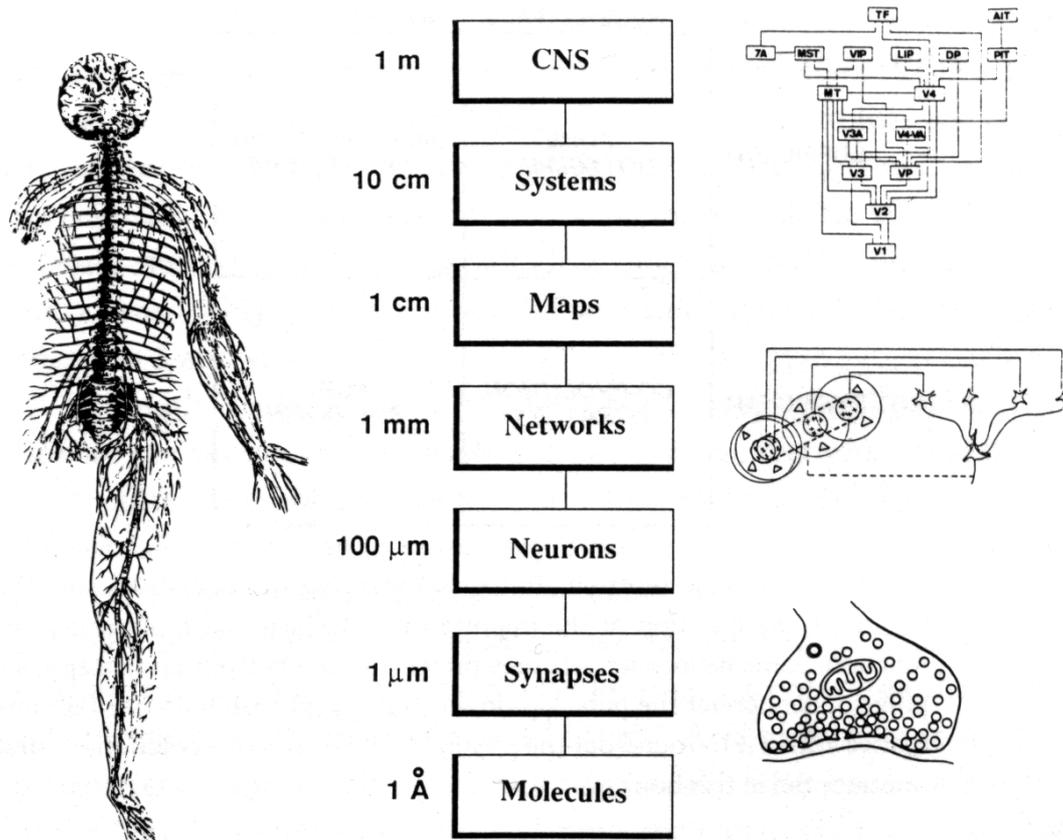


Figure 2: Churchland & Sejnowski's (1992) diagram of levels of organization in the central nervous system.

The multilevel structure of contemporary science emphasized in nearly every corner of the special sciences is not best understood as a hierarchy of theories. Nor is it a hierarchy of fields. Instead, there is an ontological hierarchy working behind the scenes. This background ontological assumption guides the development of theories, informs the criteria for evaluating explanations, and underlies the roughly hewn idea that sciences and theories are organized into levels. It is the expression of an ideal of explanation to understand how things work in terms of their component parts and to understand how those parts work in terms of still lower-level components. It is precisely because the world is presumed to have this kind of multi-level structure, of mechanisms within mechanisms, that the sciences investigating that world and the theories describing it are so reticulate that they can look like the “bio-psychological thicket” on the right side of Wimsatt’s tree. In the thicket, the orderly relationship among

levels breaks down and is replaced by a jumble. The image makes it hard to see any meaningful sense in which distinct items are at different levels. Perhaps this thought fuels eliminativism about levels.

The biological sciences are undeniably thicket-like. But this sociological fact is only indirectly related to the ontic structures presumed to lie behind and scaffold the development of these theories. From now on, then, I focus on applications of the levels metaphor to the world, not to sciences or theories.

5 Size levels

One ontological application of the levels metaphor emphasizes the relative sizes of objects at different levels. The relata in size levels are objects or kinds of object, and the interlevel relationship is relative size (larger, smaller). Things in the same size range are at the same level. Churchland and Sejnowski’s classic diagram of

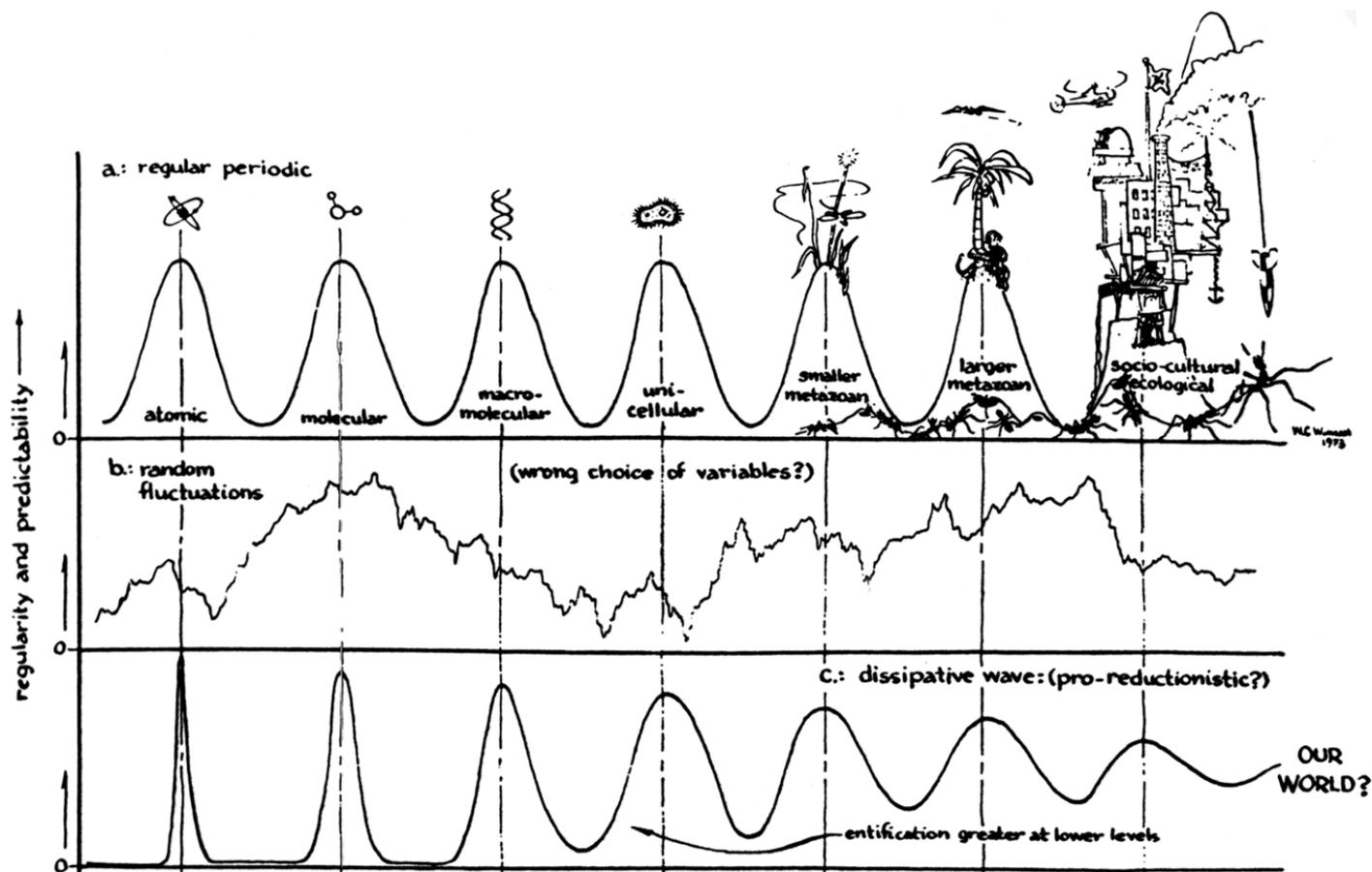


Figure 3: Levels as local maxima of regularity and at different size scales (Wimsatt 1976).

levels in the neurosciences (Figure 2) is accompanied by size scales for each level, ranging from Angstrom units to meters.⁵

As noted above, Wimsatt's tree diagram branches precisely because it is tracking something stronger than size: some kind of compositional relation. In a second diagram (Figure 3), Wimsatt emphasizes size and abandons the compositional relationship implicit in Figure 1. The abscissa in Figure 3 represents a roughly logarithmic size scale, and yet the figure is not compositional. Large metazoan organisms are not generally composed of smaller metazoan organisms, and it would surely be a stretch to claim that these two are generally composed of unicellular organisms (though there might be some truth in that claim). The ordinate in this diagram is a measure of regularity and predictability. The figure is repeated three times, each

illustrating a different way the world might be organized with respect to size. At the top is an orderly world (despite impending doom on the right). Objects become more regular and predictable in their behavior, and to the same degree, at certain size scales. Beneath this is a world with no sharp peaks of regularity and predictability. As Wimsatt notes, scientists confronted by such a world might question whether they have chosen the right variables for their models. On the bottom is Wimsatt's conjecture for our world, where regularity is very high for single atoms but falls off at larger or smaller scales. This wave dissipates over time, peaking lower and spreading out over larger and larger size ranges as scale increases. Wimsatt's diagram thus represents an *empirical hypothesis* about how levels, as peaks of regularity and predictability, are in fact distributed across different size scales in our world. If his empirical hypothesis is correct, it calls out for explanation that our world is more like the first and third graphs than it is like the second.

⁵ One might, analogously, arrange a hierarchy of activities, with different activities occurring on different temporal scales. The idea of levels of mechanisms combines these two ideas; it is a hierarchy of doings framed by a relevance relationship between those at lower levels and those at higher levels.

Why does Wimsatt represent size as the determining factor in regularity and predictability? The answer turns on other features in his prototype account. For example, things of different sizes effect and are affected by different forces, and objects of different sizes act and interact with one another more than they interact with objects at other levels. Market forces run economies, cosmic objects move under gravitational forces, and hydrogen bonds hold molecules together. Regularity and predictability peak at different size scales because the forces act and the causal relationships occur mostly at those size scales.

Wimsatt's empirical hypothesis has not been tested.⁶ Despite its intuitive appeal, one can readily produce examples of causes, forces, and laws that operate promiscuously across a very wide range of size scales. Big things (even very big things) and little things (even very little things) routinely interact, as when planets attract molecules into atmospheres or when a five-millimeter louse attaches itself to a thirteen-meter gray whale. Forces also act at many scales. Gravitation affects the human species on an evolutionary scale just as much as it influences individual human actions and the otoliths in our vestibular system. The very existence of interlevel theories, bridging molecules to behaviors (for example), provides ample evidence that regularity and predictability often span size scales: facts about gasses can be predicted from facts about molecules, and facts about learning can be inferred from facts about molecules.

If we could find a way to test Wimsatt's hypothesis, it might turn out that causes, forces, and laws do tend to cluster around certain size scales. This would be a striking empirical fact about the world and would, again, call out for some kind of explanation. In contrast, Wimsatt raises no *principled* objections to interlevel causes, forces, or regularities; he offers an *empirical* hypothesis that interlevel causes, forces, and regularities tend to be less prevalent than those operating at a single level.

⁶ It is hard to say how it would be tested and, in particular, how predictability is to be measured. Surely items in the valleys of this diagram are not unpredictable, full stop. Rather, they are more difficult to predict for creatures *like us*, unaided by machines and programs. It is not clear why human cognitive abilities should have any further ontological significance.

There are, however, apparent principled, *conceptual* difficulties faced by the effort to describe *levels of realization* in terms of a causal relation. There are many notions of realization, often tailored to altogether distinct philosophical disputes (Craver & Wilson 2007). On most accounts, however, one and the same object or event has both the realized and the realizing property, and the object cannot differ with respect to the realized property without the realizing property being different in some way (supervenience). The relata here are properties. The interlevel relationship is or includes supervenience.⁷

Marr's levels, as I understand them, are levels of realization. The hardware realizes the algorithm, which, in the right context, realizes the computation. It is awkward at best to say that the algorithm causes the computation; rather, the algorithm implements the computation in context. Changing context can change the computation. For example, a subtraction algorithm can implement division; the log of a division is a difference of logs. Likewise, the function represented in the algorithm is not caused by the hardware; the hardware instantiates or implements the algorithm. Computation-, algorithm-, and hardware-level theories are all different ways of describing one and the same thing—different predicates applied to one and the same system as a whole in its working context.

The same holds for what we might call micro-realization: when some property of a whole is realized by the organized and interacting collection of parts that constitute the property of the whole. An early edition of the *Betty Crocker Cookbook* apparently contains an explanation of how the microwave heats the soup (Churchland 1995). According to this explanation, the molecules excited by the microwave rub against one another and heat the soup by friction. As Churchland points out, Betty misrepresents the relationship between the heat of a liquid and the kinetic energy of its constituent molecules. Temperature is not produced by the

⁷ Supplemental conditions might be added to make realization more demanding (e.g., Melnyk 2003; Haug forthcoming). The point I wish to make doesn't turn on this matter.

mean kinetic energy of component molecules in such cases; rather, temperature in such situations is constituted or realized by (Churchland would say identical to) the mean kinetic energy of the components. In the same way, one might think that the behavior of a mechanism as a whole is realized by, rather than caused by, the organized collection of its components. The beating of the heart is realized, not caused, by the choreographed movements of the auricles and ventricles. It is awkward and unnatural to assert otherwise.

The apparent awkwardness and unnaturalness of such ways of talking follows from many core principles that many (rightly or wrongly) embrace about the nature of causation. If one thinks that causes must precede their effects, and one understands the realization relationship as a synchronic relation, then levels of realization cannot be causally related. If one thinks of causation in terms of the intersection of processes and the exchange of marks or conserved quantities, then the relata in levels of realization do not *come to* intersect in space-time (they always and everywhere intersect), they do not carry their marks beyond the locus of the intersection (because they always and everywhere intersect), and they do not pass anything from one to the other. In short, the *intimacy* among levels of realization seemingly precludes any standard metaphor of production, or “oomph,” or expression of a disposition, or the exertion of a power. This intimacy stands in the way of anyone who believes that causes and effects must be altogether distinct from one another.⁸ So indistinct are levels of realization that many philosophers, Churchland included, prefer to speak of identity in such contexts (see Polger 2006). Finally, if one thinks of causation in terms of the ability to manipulate effects by intervening on causes, one will note that there is no way to intervene to change the properties of wholes without, at the same time, intervening to change the supervenience base of those properties.⁹

⁸ I discuss a representative quote from Lewis below when considering causal relations between levels of mechanisms.

⁹ One can, in cases of multiple realization, intervene into the parts and their organization without intervening to change the property of the whole, and this affords some measure of independence. Perhaps one can find room in this view for the idea of understanding bottom-up relations in a hierarchy of realization as causal (though, again, realization or token identity seem to be better ways of talking). But there

I raise these issues not to cement a case against the possibility of understanding realization and causation so as to leave conceptual space for causation between levels of realization. (For a fuller discussion, see Kim 2000; Craver & Bechtel 2006). I mention them only to point out that relations of size and realization have very different implications for the intelligibility of interlevel causation. No theory or principle of causation that I know places any metaphysical restrictions on causal relations among objects of different sizes. Many theories or principles of causation appear to rule out the possibility of causal relationships between levels of realization. The point is that Wimsatt’s empirical hypothesis that causes, laws, and regularities tend to be sequestered within size scales is altogether distinct from the claim that there is no conceptual room for causation between levels of realization. Interlevel causation is mysterious or not depending on which views of levels and causation one adopts.¹⁰

6 Parts and wholes

A distinct and indirectly related application of the levels metaphor in the neighborhood of

is no room in the view (no conceptual room) for causation to work from the top down in such levels. For a penetrating discussion of this matter and its implications for causation in a multilevel world, see Baumgartner 2010, 2013; Romero (forthcoming).

¹⁰ Levels of control and levels of processing, in contrast, are defined in terms of causal relations. In *levels of control*, the relata are agencies and the relation is dominance. Items at higher levels direct or regulate the activities of their underlings. Majors and corporals, queen bees and drones, bosses and workers occupy different levels of a control hierarchy. Analogous relations are sometimes found among physiological systems. When one speaks of “executive function” in cognition, one is describing levels of control.

The idea of control or dominance is a causal notion, and it is independent of matters of size (witness the sauropod brain). Contra Fehr (2004), the idea that the world is organized in levels of realization or organization (as defined below) is not an expression of patriarchy; it is an equivocation to characterize realization and organization as relations of dominance. In levels of control, the relata are logically independent and spatiotemporally distinct interactors. It is not at all implausible for one to control the other causally (more on this below).

In *levels of processing*, the relata are processing units of some sort (such as brain regions or computational modules), and they are related as “upstream” or “downstream” in the flow of information or the order of production. In the early visual system (neglecting feedback for the moment), one can describe visual information as passing from lowest- (shallowest-, earliest-) level processing in the retina to higher- (deeper-, later-) level processing in the Lateral geniculate nucleus (LGN) and the primary visual cortex. Levels of realization and organization are not earlier or later than one another. Craik and Lockhart define levels of processing in terms of depth of semantic or cognitive processing, not in terms of decomposition.

levels of realization invites a different kind of equivocation, this time concerning the existence of higher-level powers. This application involves not a whole-whole relationship but rather the relationship between the behavior or property of a whole and the behaviors or properties of one of its parts. The behavior of the whole does not (except in special cases) supervene on the operation of the individual parts. The grain of sand contributes to the mass of the sand pile. The kidney contributes to the capacity of creatures to maintain plasma osmolality. In each case, the property of the whole (the mass of the pile, the regulation of plasma osmolality) might differ even when the contribution of these singular parts remains the same. In *part-whole* levels, as opposed to levels of realization, the relationship is between parts and wholes, not between wholes and the corresponding organized collections of entities, properties, and activities. In this case, eliminativism about levels is a non-starter (whatever its metaphysical credentials); it is impossible to imagine neuroscience, biology generally, or indeed most special sciences without the idea that things have parts.

In applying the levels metaphor to this part-whole relation, one emphasizes the relations question over the placement question.¹¹ In levels of size, things are at a given level because they are similar in size. Levels, thus conceived, are *monolithic*: they reach across all of nature, embracing everything within a given size range. Oppenheim and Putnam's layered model of the world might be read as similarly monolithic. Wimsatt's tree diagram breaks with this monolithic view precisely because it emphasizes compositional relationships: different branching levels are required because different kinds of whole (cosmological objects, human societies) are composed in different ways. If one centers part-whole thinking in one's application of the levels metaphor, then the metaphor carries no particularly useful answer to the placement

¹¹ I am here using the terms "part" and "whole" in an intuitive and inclusive way. Much of the literature on the metaphysics of parthood is simply unrelated to the many senses of part and whole used in the theories of the special sciences. I am not thinking only of objects or sets, but also about events and temporal units. I sketch a more restrictive kind of part-whole relation below, but this remains an open question (see Sanford 1993).

question. One can offer only a necessary condition: two things are at the same level only if they are not related as part to whole. Given that most things are not related to one another as part to whole, the resulting idea of being "at" a part-whole level has little or no conceptual significance.

Levels of parts and wholes lack many of the features in Wimsatt's prototype of levels. Many of the features in that prototype appear to derive from the monolithic conception shown in Figure 3: causes, forces, and laws are most plausibly thought to cluster together on the assumption that size is relevant to which forces can act, that causal relations are expressions of forces acting, and that laws govern these interactions. But if one places the part-whole relation in the center of one's metaphor, then there is no reason to embrace an empirical association between being at a given level and having a proprietary set of causal interactions for that level. Levels of parts and wholes must also be correlated with size differences because parts can be no larger than the wholes they compose. But the size differences between levels of parts and wholes are an accidental consequence of the part-whole relationship itself, not part of defining what it is for things to be at different part-whole levels.

In the following subsections, my goal is to sketch some contours of the relevant notion of part and whole. I start with classical mereology only to make the point that this apparatus was not constructed with an eye to developing a descriptively adequate account of the levels described by science. The more we learn about the limits of these classical models for our present purposes, the more we place constraints on the relevant notion of levels that, as I and others have argued, is central to the explanatory structure of neuroscience and the special sciences generally: *levels of mechanisms* (Bechtel 1988; Bechtel & Richardson 1993; Craver 2001; Machamer et al. 2000).

6.1 Types and tokens of parts and wholes

The Gibson SG has two humbucker pickups. My Gibson SG has two humbucker pickups. Not all

Gibson SG's have two humbucker pickups. But that's the factory model, the central exemplar or prototype against which variations are evaluated as more or less "typical." Likewise, when we talk about *the* human brain or *the* frog kidney, we are talking about types: the human type of brain, the frog type of kidney. And we talk also about the parts these types of things typically have.

The monolithic, layer-cake image in the Oppenheim-Putnam hierarchy is a mereology of types: societies are formed of organisms, organisms of cells, cells of molecules, and so on. Wimsatt's tree also represents relationships between types. Crystals are made of crystalline solids, which are made of inorganic molecules, which are made of atoms. It is true, of course, that all cells are made of molecules and that all organs are made of cells. But is not true that all cells are at a higher level than molecules generally.

It seems natural and harmless enough to treat the part-whole relations among types as generalizations over relations between particular wholes and particular parts. When one says, "The human brain has two hemispheres and a corpus callosum," one asserts that having these parts is typical of human brains. One is warranted on the basis of such a claim (though not always correct) in asserting of a particular human brain that it has two hemispheres. That is, the relation among part and whole *types* derives from a more primitive, token relationship between particular parts and wholes. Type-level claims about part-whole relations assert that such part-whole relations regularly or typically hold in the individuals in the relevant reference class.

One of the many useful insights contained implicitly in the branching structure of Wimsatt's tree diagram (Figure 1) is that different types of wholes are made up of different types of parts and are naturally decomposed into different levels. Both the human brain and the frog kidney are organs, and both are made of cells, but the cells in each are different, and these cells are organized differently into higher-level components. If we look within the human brain, we find that different brain regions are composed of altogether different components

and exhibit more or less proprietary organization. Broadman mapped the brain by studying these differences in cytoarchitecture from one brain region to the next. The receptive field organization of the visual cortex has, to my knowledge, no companion in the organization of the amygdala or of the mammillary bodies. Different types of brain regions/systems are made up of different types of components: they have different part-whole levels.¹²

When we say that objects or processes of one type are parts of objects or processes of another type, we are asserting that *ceteris paribus* token objects of the one type are composed of token objects of the second type. Indeed, standard attempts to define the part-whole relationship with logical rigor are expressed in terms of relationships among token individuals (Varzi 2014).¹³

6.2 Mereology

Although Oppenheim and Putnam describe the layer-cake structure of science in terms of different types of objects, the mereological structure they use to support this picture is expressed in terms of tokens. Classical mereology provides a very general and content-neutral account of the part relation. It can be used equally well to express the sense in which

¹² Elephant hearts are parts of elephants, and puffin hearts are parts of puffins, and hearts are parts of organisms. Yet puffin hearts are not at a lower part-whole level than elephants, and elephant hearts are not at a lower part-whole level than puffins.

¹³ I shall not enter here into the difficult question of how token properties, processes, and objects are individuated in the biological sciences. Consider the spatial memory system in a rat. If we take this to be one thing over the life of the organism, then it will be composed of many different sets of parts over the course of its existence, like the ship of Theseus. If we take it instead to be the spatial memory system involved in learning the layout of one particular maze, which learning might be constituted by multiple trials and extended investigation, then again we will have a single higher-level system composed differently at different time slices. If we focus on a given instant in time, then no learning can occur; learning is a kind of change. For now, I simply note (along with Bechtel & Mundale 1999) that the appropriate mapping between such parts and wholes presupposes a criterion of individuation for the whole, and what counts or does not count as a part will be determined by whether it contributes to that whole, however specified. This is related to Marcus' (2006) thought that any token identity between levels presupposes a (non-dummy) sortal that fixes the individuation conditions of the relata in the same way. Carl Gillett (2013) has called attention to the need for different accounts of compositional relations for properties, processes, and objects. Here, I am glossing over these differences to keep the discussion simple.

the word apple has the letter “a” as a part, in which courage has judgment as a part, and in which apple pie has cinnamon as a part. Common axioms associated with classical mereology, including the mereology adopted by Oppenheim and Putnam (Rescher & Oppenheim 1955), include:

- 1) **Reflexivity:** Every object is a part of itself.
- 2) **Transitivity:** Every part of a part of an object is part of the object.
- 3) **Extensionality:** An object is completely determined by the set of its parts; i.e., for objects to be identical, it suffices that they have all and only the same parts.
- 4) **Summation:** Any pair of objects (x , y), is itself an object, z , which is their sum.

I list these constraints only to illustrate that classical mereology will take us only so far in the effort to characterize part-whole levels. This is a formal theory, abstracted entirely from the concerns of practicing scientists. This means that there are constraints in the scientific conception of parts and wholes that classical mereology need not honor. First, the levels that Churchland and Wimsatt describe are space-, structure-, and time-involving in ways that classical mereology need not be. The set of integers is part of the set of real numbers, and “Consider the Lobster” is part of Wallace’s corpus, but not in the same way that the glutamate receptor is part of the chemical synapse. The glutamate receptor takes up part of the space occupied by the synapse as a whole. Its opening is part of the extended process by which neurons communicate. None of this is expressed or intended to be expressed in the generic part-whole relation of classical mereology.¹⁴

¹⁴ Marr’s levels are not space-involving in this way. The algorithm is not located within the computation, it is not a substage of the computation, and it is not organized together with other parts in the service of the computation.

Second, although reflexivity is involved in certain theoretical applications of mereology, it has no application in thinking about such space-involving levels. It does no justice to the biological concept to assert that every hippocampus is a part of itself. If levels are defined as a relationship between a part and a whole, and everything is a part of itself, then everything is at both higher and lower levels than itself. The parts surely must be proper parts.¹⁵

Third, it has been noted that the transitivity axiom often fails to apply to functional parts of the sort that populate physiological and biological theories (Varzi 2014). Eric is part of the championship pool team, and Eric’s locks are part of Eric, but his locks are not part of the team. However, if one requires of a part (entity, property, or activity) that it must be *relevant* to the property or behavior of the whole, then one can retain the transitivity of this relation, at least in many contexts. *If we ask not about Eric, but about the motion of his arm as he wields his cue*, then his locks are clearly not relevant while the muscles gliding his arm steadily forward *are* relevant. So too are, in some sense, the molecules transmitted across the neuromuscular junction during his backstroke. When what counts as a part is filtered in each iteration by explanatory relevance relations (they are not mere spatial or temporal parts but working parts—parts that are involved in, contribute to, or make a difference to the property or activity of the whole), then the relationship is, in fact, transitive. The appearance of a failure of transitivity in functional systems trades, it seems, on failing to relativize the decomposition into parts by a highest-level target (explanandum) phenomenon; not all the spatiotemporal parts of an object or process are relevant to everything it does. It is only relative to a highest-level activity or property of the hierarchy under consideration that the lower level parts are visible as components—as working parts in the mechanism. *If we think not about the team, Eric, and his locks, but rather*

¹⁵ This is not a big departure from classical mereology; one could simply restrict one’s attention to proper parts. But the point underscores the fact that classical mereology was not developed with an eye toward understanding the sense in which pyramidal cells are parts of the hippocampus.

about the victory, the shot, and the muscular contraction, matters seem different.¹⁶ The contraction contributes to the shot, which contributes to the victory. The locks will not appear in this hierarchy, but the relevant parts will.¹⁷

The extensionality theorem holds that no two distinct objects share all and only their proper parts. A hippocampus and a bust of the Dalai Lama formed out of the same pyramidal cells, granule cells, etc. that compose the hippocampus are, according to classical mereology, one and the same object. But in biological systems, the organization of components is often (perhaps always) relevant to the properties and activities of the whole. Again, parts appear as parts only relative to a decomposition framed by reference to some highest-level property or activity. This is Kauffman's point, enshrined in Glennan's law: a mechanism is always a mechanism of a given phenomenon (Kauffman 1971; Glennan 1996). Thus Kauffman:

A view of what the system is doing sets the explanandum and also supplies criteria by which to decide whether or not a proposed portion of the system with some of its causal consequences is to count as a part and process of the system. Specifically, a proposed part will count as a part of the system if it, together with some of its causal consequences, will fit together with the other proposed parts and processes to cause the system to behave as described. (1971, p. 260)

The more general point is that there is an application of the levels metaphor that is not merely a part-whole relationship as specified in classical mereology, but one in which the parts are relevant (explanatorily and constitutively) to some property or activity of the whole.

One can make a similar point with respect to the summation axiom. This theorem allows

¹⁶ Clearly Eric's muscles are not part of the team, but this reflects only the fact that teams can have only certain kinds of part as members. If we look rather at an activity of the whole and ask what contributes to that, a different picture emerges.

¹⁷ Specific details about lower-level parts might be screened off in cases of multiple realization. Specific details about the parts might not be relevant. In that case, it would appear one must appeal to more abstract properties of the parts.

one to form arbitrarily many gerrymandered wholes out of disparate and unconnected parts with no spatial, temporal, causal, or functional unity. Lewis (1991) calls this "unrestricted composition": whenever there are some things, there is also a fusion of those things. The Yankees's starting rotation and the now disparate parts of my mother's old Chevy Vega together form a whole. This way of thinking about parts and wholes has little or no application in biology because such gerrymandered wholes don't do anything interesting (though such wholes will have aggregate properties of the sort discussed below). The whole in such gerrymandered collections typically doesn't play any explanatory role. And what goes for wholes goes for parts as well. According to this classical picture, it is perfectly legitimate to claim that my dog, Spike, has four parts: a front quarter, a hindquarter, and two midsections of approximately equal length. There is nothing to prevent this way of talking; but the parts revealed in this decomposition do not cut Spike at his joints. The biological decomposition finds joints at causal interfaces, and identifies parts with more or less isolable (nearly decomposable) subsystems (Simon 1962) that contribute to the behavior of the whole.

In short, many of the ideas central to classical mereology must be amended or restricted if they are to apply to the part-whole levels distinctive of biology, neuroscience, and the special sciences generally. At least some of the work can be done by restricting the part-whole relation by a *relevance condition* on biological part-whole: all the lower-level properties, activities, and organizational features of the parts are relevant to—contribute to—the property or activity of the whole.

6.3 Levels of organization: Aggregates and mechanisms

So let us focus on an application of the levels metaphor that is a part-whole relation and a (constitutive) relevance relation. I will not dwell here on the appropriate notion of relevance (see Craver 2007, see also Harinen 2014). For now, we can work with the idea that each part in

such a hierarchy (in addition to being spatially and/or temporally contained within the whole) plays a necessary but insufficient role within a collection of parts that are jointly sufficient (but possibly redundant) for a given explanandum phenomenon (Couch 2011). That is, relevant parts might usefully be thought of as constitutive insufficient, but necessary part of an unnecessary but sufficient condition (I) for the behavior of the mechanism as a whole (Mackie 1973).¹⁸

Again following Wimsatt (1997), we can distinguish two ways that spatiotemporal parts contribute to a property or activity of a whole: *aggregation* and *organization*. An aggregate property is literally a sum of the properties or activities of the parts. The current flowing through an ion channel, for example, is a sum of the charges carried by individual ions. The concentration of a volume of a fluid is a sum of the number of particles in that unit volume. Aggregative properties change linearly with the addition and removal of parts. And aggregative properties do not change as the parts are inter-substituted with one another. Some properties of the hippocampus, such as its mass, remain the same when the cellular constituents of the hippocampus are reorganized to represent His Holiness. Other properties of the hippocampus, such as its information processing capacities, are destroyed. For truly aggregative properties, spatial, temporal, and causal organization among the components is irrelevant.

¹⁸ Once we have made this adjustment, the *relata* in this relevance-merology are no longer objects but rather properties, activities/processes, or (as is more common in philosophical parlance) events. One does not explain the elephant; one explains why the elephant has large ears or how the elephant circulates its blood. One does not explain gasses; one explains their temperature and pressure. This point marks a significant departure from the mereological views of levels discussed above. Each of those applications of the levels metaphor focuses on objects or types of objects (societies, organisms, cells, and so on) as the *relata*, not on their properties, activities, and aspects of their organization. In many cases, the components picked out in a mechanistic decomposition fail to correspond to paradigmatic objects with clear spatial boundaries. The synapse, for instance, is composed of part of a presynaptic cell (the axon terminal), part of a postsynaptic cell (the dendrite or bouton), and a gap between them. What unifies these items into an object is their organized behavior: the pre-synaptic cell releases transmitters that traverse the cleft and act on the postsynaptic cell. Synapses are not cells or parts of cells, nor are they composed of cells. Rather, they are objects unified by their relevance to a given activity of the whole, such as chemical transmission.

Aggregates are rare. The masses of the individual grains in a sand pile do, in fact, depend on the spatial distribution of the other grains (if one takes relativity seriously). What is presumed to be a homogeneous concentration of a liquid can in fact have local concentration differences depending on how the ions are organized in different parts of the fluid. In the case of non-aggregates, the activity or property of the whole is not a simple sum of the properties of the individuals. Adding or removing parts (e.g., the human heart) can lead to dramatic changes in how the system (e.g., the body) works. And rearranging the parts and their activities in space and time can eliminate the explanandum phenomenon entirely (as would happen if one randomly swapped parts of the circulatory system for one another). This is all true because spatial, temporal, and causal organization are relevant to (make a difference to, partly constitute) the property of the whole.

I use the term “mechanism” permissively to describe non-aggregative compositional systems in which the parts interact and collectively realize the behavior or property of the whole. Mechanisms are by definition more than the sums of their parts: they have properties their parts do not have, and they engage in activities that their parts cannot accomplish on their own.

Most mechanisms with which I am familiar involve myriad part-whole relations, some of which are more aggregative in nature, and some of which are less so. Many things brains do, for example, involve the flux of ions across a membrane, which flux is closer to the aggregative than the mechanistic end of the organizational spectrum. Other things brains do (such as the developing grid cells in the entorhinal cortex) require precisely organized relations among the activities of cells in and around the entorhinal cortex. This organizational spectrum from aggregate to mechanism covers all the relations that go into levels of organization, the superordinate class.¹⁹

In levels of mechanisms, the *relata* are some activity or property of a mechanism as a

¹⁹ Levy (2013) calls attention to the fact that biological systems typically involve both aggregation and organization.

whole,²⁰ and the activities, properties, or organizational features of its components (its relevant parts and organization). Some component, X's φ -ing, is at a lower mechanistic level than S's ψ -ing if and only if X's φ -ing is a component in S's ψ -ing, that is, if and only if X's φ -ing is a relevant spatiotemporal part of S's ψ -ing. In levels of mechanisms (as opposed to aggregates) lower-level components are organized together to make up some behavior or property of the whole; in aggregates, the properties of the parts are summed.

Levels of mechanisms are represented in Figure 4. At the top is the activity of some mechanism as a whole (S's ψ -ing). S's ψ -ing is a behaving mechanism. Although one can speak of the mechanism and its activity separately (as when a mechanism stands inactive but ready to act), such separation in thought is artificial. Even the static mechanism is defined and sub-divided by reference to what it does. ψ is the topping-off activity of the mechanism for which all lower-level components are relevant. It can be idealized as an input-output relation, though this is an impoverished way of understanding phenomena (see Craver 2007). One level down are the activities and components, the X's φ -ing, which compose and are organized together to constitute S's ψ -ing.²¹ Below that is another iteration of levels: the ρ -ings of Ps organized such that one of the Xs φ s as it does. By organization, I mean that the parts have spatial (e.g., location, size, shape, and motion), temporal (e.g., order, rate, and duration), and active (e.g., feedback or other motifs of organization;

see Levy & Bechtel 2014) relations with one another by which they work together to do something they cannot do on their own. As noted above, the relationship between levels is a part-whole relationship filtered further by constitutive relevance (Craver 2005; Harinen forthcoming). In levels of mechanisms, parts are made into higher-level components by being organized spatially, temporally, and actively into something. In more aggregate compositional relationships, they are summed into higher levels.

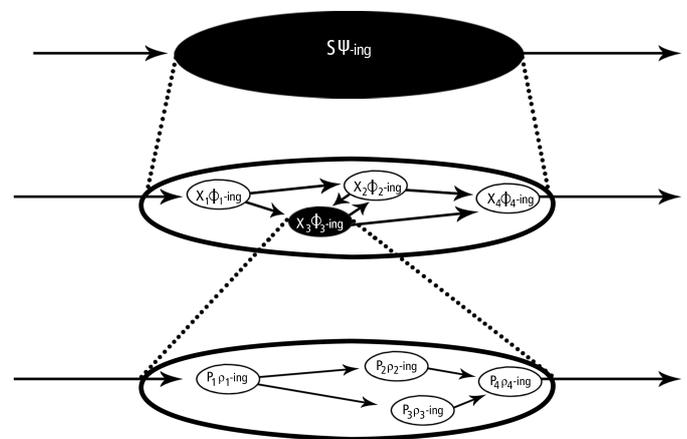


Figure 4: An abstract diagram of levels of mechanisms.

Contemporary theories of learning and memory provide a compelling example of levels of mechanisms (see Craver & Darden 2001; Craver 2002). The top level is a mechanism as a whole engaged in a spatial memory task, such as learning to run efficiently through a maze. One component in that mechanism, and so one level down in this description, is the hippocampus, a region of the brain thought to form a “map” of locations and orientations within the maze. The capacity of the hippocampus to acquire such an internal map of local spaces is thought to be explained, in part, by changes in synapses between pyramidal cells, specifically by a process known as Long-Term Potentiation (LTP). And it is now known that n-methyl d-aspartate (NMDA) receptors (n-methyl d-aspartate is a pharmacological agonist that binds these receptors preferentially), contribute to LTP. This story could continue downward, looking

²⁰ These might be understood as the obtaining of a property or the unfolding of a process over time. What counts as a static property often depends on one's temporal resolution.

²¹ I have not always chosen my language in a way that comports with common usage among metaphysicians, preferring to follow Salmon (1984). In this paper, I have tried to make it clear that I am interested in components. Components, as the name suggests, compose behavior of the higher-level mechanism when organized together. All of the component entities and activities organized together, it now seems appropriate to say, jointly constitute the behavior of the whole. That is, I am now using componency to talk about relationships between wholes and parts, and I am using “constitution” to talk about levels of micro-realization. I am not especially interested in the relationship between statues and lumps of clay. I am interested in how parts are organized and interact so that together they exhibit higher-level behaviors. I know of no metaphysician who has developed an adequate notion to express this, so perhaps I will be forgiven for appropriating these words for new uses.

into aspects of protein chemistry and the structural changes thought to underlie channel functioning.

6.3.1 Levels of mechanisms are local

Levels of mechanisms are of entirely local significance. The levels in our example are defined by reference to a topping-off point, spatial memory, contribution to which determines whether or not a spatiotemporal part of the system is in fact relevant—whether it is a component in the mechanism *for S's Ψ -ing*. The hierarchy in Figure 4 and the levels of spatial memory as I have described them follow only a single (local) strand of embedding: from the behavior of the mechanism as a whole, to the behaviors of its components, on to the behaviors of *one* of these components, and so on.

Levels of mechanisms, like part-whole levels generally, are not monolithic divisions in the furniture of the world. Levels of mechanisms are defined only within a given part-whole hierarchy. There are different levels of mechanisms in the spatial memory system, in the circulatory system, in the osmoregulatory system, and in the visual system; the levels in each need not map onto one another. How many levels there are, and which levels are included, must be determined on a case-by-case basis by discovering which sorts of components are explanatorily relevant for a given phenomenon. Levels of mechanisms cannot be read off a menu of levels in advance.

If we apply the levels metaphor only locally, then it makes no sense to ask whether the hippocampus is at a higher or lower level than the nephra in the kidney. The nephra are not part of the hippocampus, and they are not relevant to the functioning of the hippocampus. Neither similarities of size nor similarities in kinds of part are definitive of levels of mechanisms. Rather, levels of mechanisms are defined relative to one another within a hierarchically organized mechanism.

The idea that levels of mechanisms retain some hint of the layer-cake model can sneak its way back into one's application of the metaphor

if one slides unknowingly between tokens and types of parts and wholes. Compare the following three sentences:

- a) This pyramidal cell is at a lower level of mechanisms than this hippocampus.
- b) Pyramidal cells are at a lower level of mechanisms than hippocampi.
- c) Cells are at a lower level of mechanisms than organs.

Statement (a) expresses a mechanistic notion of levels: a particular pyramidal cell is a component of a particular hippocampal mechanism. This statement is true if the cell is a component in a mechanism for a given activity in which the hippocampus is engaged. It might be, for example, that a given pyramidal cell is a component in some hippocampal mechanisms but not others; if so, it is at a lower level to some hippocampal activities and not others.

Wimsatt describes the compositional relationship between levels as a relationship between types. He writes: "Intuitively, one thing is at a higher level than something else if things of the first type are composed of things of the second type" (Wimsatt 1976, p. 215). This is a departure from the idea of levels of mechanisms and one that threatens to reinstate something like the Oppenheim and Putnam hierarchy. Pyramidal cells are found outside the hippocampus, and those pyramidal cells are not parts in hippocampal mechanisms; they are not at a lower level of mechanistic organization. Likewise, both the hippocampus and the kidney are composed of cells; organs tend to be composed of cells. But the cells in the hippocampus are not at a lower mechanistic level than kidneys because they do not contribute to kidney function. The slide from sentences such as (a) to sentences such as (b) and (c) is a slide back toward the layer-cake model. Of course, scientists typically deal with types. But as I suggest above, this is a generalization over a relationship between tokens. The correct generalization is that the cells that compose hippocampi are at a lower level than the hippocampi they compose.

This is significant for two reasons. First, it helps to show that many objections to thinking about neuroscience and other special sciences in terms of levels simply do not apply to this restricted application of the metaphor. If one thinks, with Wimsatt, me, and probably Oppenheim and Putnam, that the Oppenheim and Putnam layer cake is an overly simplistic representation of the diverse ontological structures one finds in the special sciences—that things like ocular dominance columns and synapses don't readily fit that picture—one can nonetheless retain the idea that mechanisms are susceptible to multiple nested decompositions. These are different applications of the levels metaphor. Secondly, and perhaps more importantly, the idea that levels are local significantly shifts the reductionist world-view for which Oppenheim and Putnam developed their ontology of levels. If one thinks of levels as levels of organization (as levels of mechanism and levels of aggregation), then it is inaccurate to think of reduction as involving relationships among theories developed to describe the items at a particular monolithic level. If reduction is simply a matter of explaining a higher-level phenomenon in terms of the organized activities of components, then reduction is still possible within a mechanistic world-picture, but it will be achieved not through grand reductions of overarching theories, but rather through piecemeal explanatory achievements for specific phenomena. Visions of the unity of science through interlevel reduction have to be revised not as grand unifications across the whole of science but rather as local explanatory successes. Such local explanations will, in fact, integrate findings from different sciences and bring different theoretical vocabularies into conversation with one another (see Craver 2005; Craver & Darden 2001), but it only deceptively resembles the layer cake that Oppenheim and Putnam sketched as a working hypothesis.²²

²² Nothing in this picture is meant to deny token identity between the behavior of a mechanism as a whole and the organized behavior of its parts. Because there are some conceptual difficulties that stand in the way of speaking meaningfully about token identities between levels, I have written with fewer commitments about constitution.

6.3.2 Placement is weak and derivative in levels of mechanisms

One consequence of the mechanistic application of the levels metaphor is that there is no unique answer to the question of when two items are at the same mechanistic level. Only a partial answer is available: X's ϕ -ing and S's ψ -ing are at the same level of mechanisms only if X's ϕ -ing and S's ψ -ing are components in the same mechanism, X's ϕ -ing is not a component in S's ψ -ing, and S's ψ -ing is not a component in X's ϕ -ing.²³ Unlike size levels or levels defined in terms of the types of objects found at a given level, levels of mechanisms are defined fundamentally by the relations question: by the componentency relationship between things at higher and lower levels. If two things are not related as part to whole, they are not at different levels, and so, if they are in the same mechanism, they are, in this very weak sense, at the same level. But this is just to say that sameness of level has no significance within this application of the metaphor.²⁴

If one thinks of levels of organization as levels of aggregates and levels of mechanisms, then spatial containment and size relations between levels follow as an accidental consequence of the componentency relation. The pyramidal cells are contained within the hippocampus, which is contained within the spatial memory system. The activities of these entities are also related as temporal part to whole: the binding of glutamate is a

²³ This has struck some readers as circular because it appears to state that X and S are at the same level if they are not at different levels. Appearances to the contrary, this is not circular. I have defined "same level" in terms of the notion of "different level" and the latter is defined in terms of componentency relations. The appearance of circularity, I believe, results from the fact that most people assume that the notion of "same level" must be primitive relative to the notion of "different level," and I have reversed that assumed order.

²⁴ Another way to see that levels of mechanisms do not answer the placement question is to recognize an apparent failure of transitivity. Suppose X1 and X2 are components in the same mechanism, that neither is a component in the other, and that the behavior of X2 can be decomposed into a set of interacting components, including P1. X1 would, according to this account, be "at the same level" as both X2 and P1 even though X2 and P1 are at different levels from one another. This problem, first raised by Lindley Darden (personal communication), is only a problem if one demands that there must be a unique answer to the placement question for an account of mechanistic levels. My argument against the notion of monolithic levels turns on the absence of any good principle for stretching the ideal of levels beyond its local context.

temporal component in the activity of the NMDA receptor. The objects at lower levels are smaller than (or at least no larger than) the whole, giving the hierarchy a derivative size ordering. Relations of size, rather than defining what it is for an item to be at a level (the placement question), are derivative from the more fundamental relationship between levels (the relations question): namely, the relationship between a mechanism and a component.

6.3.3 Emergence and levels of mechanisms

Mechanisms do things that their components taken individually cannot. This marks a sharp distinction between levels of mechanisms and levels of realization. Kim says this point is “obvious but important”:

This table has a mass of ten kilograms, and this property, that of having a mass of ten kilograms, represents a well-defined set of causal powers. But no micro-constituent of this table, none of its proper parts, has this property or the causal powers it represents. H₂O molecules have causal powers that no oxygen and hydrogen atoms have. A neural assembly consisting of many thousands of neurons will have properties whose causal powers go beyond the causal powers of the properties of its constituent neurons, or subassemblies, and human beings have causal powers that none of our individual organs have. Clearly then macroproperties can, and in general do, have their own causal powers, powers that go beyond the causal powers of their micro-constituents. (Kim 1998, p. 85)

Through aggregation or organization, wholes have causal powers that their parts individually do not have. An activity at a higher level of mechanistic organization is quite literally more than the sum of its parts. It is not an aggregate. The addition and removal of parts leads to nonlinear changes in the behavior of the mechanism as a whole. It matters how the parts are organized; it is in virtue of their organization that

they have properties that go beyond the properties of the individual parts (Wimsatt 1996). This feature of levels of mechanisms is so obvious, so prosaic, and so banal as to be hardly worth mentioning. No fancy complexity is required: two toothpicks stacked perpendicular to one another have the mechanistically emergent capacity to act as a lever or catapult; neither toothpick can do so on its own.

Of course, most mechanisms in biology are substantially more complicated than that. They have many more parts. Those parts interact with one another with bewildering complexity. Often they contain feedback relations that introduce nonlinear interactions into the operation of the mechanism itself. The mechanisms of LTP, for example, have yet to yield their secrets completely despite the dedicated attention of thousands of researchers over forty-odd years. A glance at any recent textbook is enough to convince one that LTP involves myriad intracellular reactions, protein synthesis, structural features of dendritic spines, changes to vesicular release, and retrograde transmission with nitric oxide. The mechanism involves so many parts and interactions that it would be useless, if not impossible, to represent them all in a visual diagram. Keeping track of how they all work together would require a very complicated computational model of some sort that has yet to be developed. As mechanisms get this complicated, we reach the limits of our ability to predict how the behavior of the whole will change as the parts change. Any change introduced to a part has so many ramifications that it is difficult or impossible for creatures like us to keep track of them all. This is an interesting fact about us and the limits of our cognitive and modeling prowess. But, ontologically, it is the same old banal fact about the importance of organization in mechanisms. We have added only that we have difficulty keeping track of it all.

Likewise, a common scientific complaint against reductionistic research programs in biology and neuroscience is that one can make only limited progress by studying the parts of mechanisms in isolation from one another.²⁵

²⁵ This idea of reduction is not the standard notion of theory reduction but something closer to what Eric Kandel means

We can study LTP in cells grown in a culture, and we can study hippocampal computations in a razor-thin hippocampal slice, and we can study spatial learning in highly contrived environmental settings such as a large pool filled with milky water (the Morris water maze). Such reductionist practices are absolutely essential to progress in the sciences. Nonetheless, one engaged in such practices must (and typically does) bear in mind that the behavior of the part when it is isolated for experimental purposes might be very different from the behavior the part exhibits when it is working in the context of a mechanism. Causal interactions with other parts of the mechanism and background conditions “in the wild” might lead to behaviors that would never be discovered in such simplified preparations. This is an extremely important point about reductionist research programs (Bechtel & Richardson 1993), and one might choose to describe this well-known difficulty with the language of emergence. But this is just to say that one cannot truly understand how a mechanism works until one understands how all its parts are organized together and working in the relevant conditions, and this we have already said repeatedly.

I emphasize the banality of these observations to stress that many of the things one wants to say about organization in biological systems can be said within the mechanistic application of the levels metaphor without introducing anything that is metaphysically novel or suspect. As the complexity of a mechanism increases, the epistemic challenges we face in discovering and modeling it increase as well, but this is of no significance for the ontic structures—the entities, activities, and organizational features that exist in the world.

Not so for *spooky emergence*. Spooky emergence is spooky precisely because it is committed to the existence of higher-level properties that have no explanation in terms of the parts, activities, and organizational features of the system in the relevant conditions. Levels of mech-

anisms are levels of ontic mechanistic explanation (Craver 2014): they are defined in terms of componency and constitutive explanatory relevance. If that explanatory relationship is severed, then the sense in which emergent properties are at a “higher level” must be altogether different than the compositional notion of levels in levels of mechanisms. If one imagines that atoms compose molecules, which are organized into cells, which are linked into networks from which mental properties spookily emerge, the first three steps are upward steps in a hierarchy of levels of mechanisms, but the last is not. The ability of organization to elicit novel causal powers (that is, nonaggregative behaviors and properties) is unmysterious both in scientific common sense and common sense proper (Van Gulick 1993; Kim 1998). Appeal to strong or spooky emergence, on the other hand, justifiably arouses suspicion. Indeed, it is unclear why properties that emerge in a spooky fashion should be thought of as higher-level at all. Perhaps the very idea of spooky emergence is incoherent.

6.3.4 Mechanistic levels are not causally related to one another

As with levels of realization, many common assumptions about the nature of causation would appear to make causal relations between mechanistic parts and the properties or behaviors of wholes suspect. Items at different levels of aggregation and at different levels of mechanisms are intimate with one another in much the same way that items at different levels of realization are intimate. Lewis is explicit. If C causes E:

C and E must be distinct events [if they are to be causally related]—and distinct not only in the sense of nonidentity but also in the sense of nonoverlap and non-implication. It won't do to say that my speaking this sentence causes my speaking this sentence; or that my speaking the whole of it causes my speaking the first half of it; or that my speaking causes my speaking it loudly, or vice versa. (Lewis 2000, p. 78)

by the term: choosing to study complex phenomena in extremely simple systems. We might call this experimental reductionism.

The relevant kind of intimacy for levels of mechanisms is overlap between token events or processes. The relationship between LTP and the opening of NMDA receptors during LTP induction is directly analogous to the relationship between speaking the whole of a sentence and speaking its first half. The induction of LTP is partly constituted by the opening of the NMDA receptor. The would-be cause in this top-down causal claim already contains the would-be effect within it. There is nothing additional to be produced in the effect; the occurrence of the effect includes the occurrence of the cause.

What about the bottom-up case? We might say that the spark plugs cause the engine to run, all the while acknowledging that the sparking of spark plugs is part of the operation of the engine. The naturalness of this locution is at least partly due to an ambiguity in the way we commonly describe the behavior of a mechanism as a whole. Sometimes we describe it as an activity or process that starts with the mechanism's setup conditions and ends with its termination conditions (Machamer et al. 2000). Thus we might describe Long-Term Potentiation as a *process*²⁶ or activity beginning with a rapid and repeated stimulus (called a tetanus) to the presynaptic neuron and ending with enhanced transmission across the synapse. Other times we describe the behavior of the mechanism as a whole, the phenomenon, as the *product* of that process (or one of its termination conditions). Thus we might say that the mechanism of Long-Term Potentiation produces a *potentiated synapse*. This way of speaking leads us to think in terms of the antecedent causes of potentiation: the tetanus is a distal cause, and the subsequent changes in the NMDA receptor are more proximal. If we think about the behavior of the mechanism in the second way, as a product, it is natural to think of the opening of the NMDA receptor as a cause of the synapses being potentiated (and indeed it is). But if we think about the behavior of the mechanism in the first way, as an input-output relation starting with the tetanus and ending with a potentiated synapse, then it is wrong to think of the tetanus or the opening of the NMDA receptor as a cause of *that*. The

NMDA receptor is a part of that causal process. These are two equally acceptable ways of describing the relationship between a mechanism and a phenomenon; they are easily translated into one another. However, if one is careless, these ways of speaking and writing invite equivocation of precisely the sort that we are struggling here to avoid.

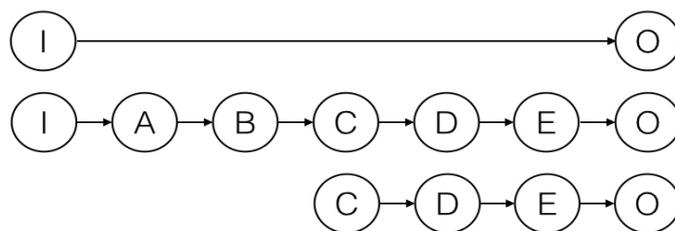


Figure 5: Why bottom-up causation is conceptually problematic.

Suppose we represent the input-output relationship constituting LTP as in the top of Figure 5, where I is a tetanus and O is a stable, potentiated synapse. Beneath this abstract I-O relation is a more detailed description of the intermediate stages in this mechanism: the tetanus excites the postsynaptic cell (A) which depolarizes it (B), causes NMDA receptors to open (C), and so on. (Nothing turns on the fact that I've idealized this mechanism as a single, step-wise causal chain.) Now, suppose we intervene to open the NMDA receptors directly and thereby potentiate the synapse (as shown in the third line). We might say that this intervention induced LTP; but when we say this, we mean that it produces the end product of the mechanism (it potentiates the synapse). We cannot coherently assert that it causes the process as depicted in the first two lines. This is for the simple reason that the process in the first two lines includes stages I, A, and B, and these are absent in the causal sequence represented in the third line. At most we can say the intervention caused the last half of the process. The NMDA receptor is not a cause of the process of LTP; it is a component of that process.²⁷

²⁷ This case is easiest to make for interventions that start the process midway. If an intervention, instead, were to augment C and thereby produce a more potentiated synapse than one would otherwise have had, then causal language would appear to be appropriate. The intervention changes, makes a difference to, the input-output relationship. These considerations generalize naturally to claims about types of

²⁶ Not in the Salmon (1984) sense, but in the colloquial sense of an unfolding sequence of states and activities.

6.4 Levels of mechanisms in relation to other kinds of levels

This application of the levels metaphor, according to which levels of organization are understood in terms of levels of aggregation and levels of mechanisms, thus offers a no-nonsense, ontological picture that comports well with the kinds of explanatory structure one finds in neuroscience and throughout the special sciences generally.

This view eschews the idea that levels are monolithic strata in the structure of the universe, with proprietary causal laws and forces (contra the view in Wimsatt, Oppenheim, and Putnam). Likewise, it allows that items at higher levels have causal powers that things at lower levels do not, in contrast to levels of realization. Single sciences and theories might investigate phenomena at many levels of organization, and an item located at one level in such a compositional hierarchy might be studied by many sciences and described in many theories. Things at different levels of organization (aggregation and mechanism) do not causally interact with one another, though we might find more complicated ways of describing how these items depend upon one another (see Craver & Bechtel 2007). As a result, if we think about the world in terms of levels of organization, we should not be tempted into thinking that things at higher levels control or dominate things at lower levels. Levels of organization are, in a sense, levels of explanation, given that explanations for different topping off phenomena will often decompose the system into altogether different parts within parts. It might be difficult to discern such levels in scientific practice, and the organization of components might be very complex, but nothing emerges from levels of mechanisms except in the banal sense that parts organized together do things that the parts alone cannot. Levels of organiz-

mechanistic parts and wholes. Separately: When we describe this relation as a kind of production, levels show up as intermediate causes. Perhaps the temptation to speak of levels at all is lessened if one maintains that perspective. But this is not a change in what is being said so much as a change in how it is being said.

ation, in other words, seem to capture many of the intuitions that accompany the idea that the world is organized into levels but without many of the objectionable elements of other applications of the levels metaphor. The fact that the levels metaphor is often used carelessly and deployed in ways that violate common sense and metaphysical ideas about the structure of the world should not lead one to abandon the metaphor entirely. As we've seen, it can be given a relatively precise and metaphysically unobjectionable formulation that, in addition, fits the multilevel structures that the most advanced special sciences seem to be discovering.

7 Conclusion

Despite the ubiquity of levels talk in contemporary science and philosophy, very little has been done to clarify the notion. Here I defend a kind of descriptive pluralism about the levels metaphor: it is applied usefully in many contexts to describe different relations, different relations, and different senses in which items might be located at a given level. Because the levels metaphor is so ubiquitous and so promiscuously applied, some vigilance is required to keep the applications distinct from one another. I have discussed only a few applications: levels of science, theory, realization, size, mereology, aggregation, and mechanism. Even in these few key examples, we have found good reason to remain vigilant. The implications of the levels metaphor in one application only occasionally transfer when the metaphor is applied differently.

I have also suggested that levels of mechanisms (or, more generally, levels of organization) are especially important to the explanatory structure of neuroscience and the special sciences generally. If one thinks of levels in this way, one can easily see why interlevel causation should seem so problematic (indeed, it is problematic), one is free to jettison Oppenheim and Putnam's idea of monolithic levels of nature, and one can see room in the causal structure of the world for the existence and legitimacy of higher-level causes and explanations. Whether

the idea of levels of mechanisms truly pays off in such useful ways remains to be seen. I merely hope to have preserved the metaphor, and its application to mechanisms, in the face of problems it inherits only through equivocation.

Acknowledgements

This essay is dedicated to William C. Wimsatt, whose sensitive and pioneering explorations of complexity, levels, and organization have shown how to tame the biopsychological thicket without sacrificing its wildness. Thanks to Anthony Dardis, Jens Harbecke, Donald Goodman-Wilson, Eric Hochstein, Eric Marcus, Lauren Olin, Gualtiero Piccinini, Anya Plutynski, Philip Robbins, Mark Povich, Felipe Romero, and Gary Williams for comments on earlier drafts.

References

- Baumgartner, M. (2010). Interventionism and epiphenomenalism. *Canadian Journal of Philosophy*, 40 (2), 359-384. [10.1353/cjp.2010.0015](https://doi.org/10.1353/cjp.2010.0015)
- (2013). Rendering interventionism and non-reductive physicalism compatible. *Dialectica*, 67 (1), 1-27. [10.1111/1746-8361.12008](https://doi.org/10.1111/1746-8361.12008)
- Bechtel, W. & Mundale, J. (1999). Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of Science*, 66 (2), 175-207. [10.1086/392683](https://doi.org/10.1086/392683)
- Bechtel, W. & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton University Press.
- Bechtel, W. (1988). *Philosophy of science: An overview for cognitive science*. Hillsdale, NJ: Erlbaum.
- (2013). Addressing the vitalist's challenge to mechanistic science: Dynamic mechanistic explanation. In S. Normandin & C. T. Wolfe (Eds.) *Vitalism and the scientific image in post-enlightenment life science 1800-2010*. Dordrecht, NL: Springer.
- Bickle, J. (1998). *Psychoneural reduction: The new wave*. Cambridge, MA: MIT Press.
- Churchland, P. S. & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, MA: MIT Press.
- Churchland, P. S. (1995). Can neurobiology teach us anything about consciousness? *Proceedings and Addresses of the American Philosophical Association*, 67 (4), 23-53. [10.2307/3130741](https://doi.org/10.2307/3130741)
- Couch, M. (2011). Mechanisms and constitutive relevance. *Synthese*, 182 (3), 119-145. [10.1007/s11229-011-9882-z](https://doi.org/10.1007/s11229-011-9882-z)
- Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, 11 (6), 671-684. [10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- Craver, C. F. & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, 22 (4), 547-563. [10.1007/s10539](https://doi.org/10.1007/s10539)
- Craver, C. F. & Darden, L. (2001). Discovering mechanisms in neurobiology: The case of spatial memory. In P. K. Machamer, R. Grush & P. McLaughlin (Eds.) *Theory and method in neuroscience* (pp. 112-137). Pittsburgh, PA: University of Pittsburgh Press.
- Craver, C. F. (2001). Role functions, mechanisms and hierarchy. *Philosophy of Science*, 68 (1), 31-55.
- (2003). The making of a memory mechanism. *Journal of the History of Biology*, 36 (1), 153-195.

- [10.1023/A:1022596107834](#)
- (2005). Beyond reduction: Mechanisms, multifield integration, and the unity of science. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36 (2), 373-396. [10.1016/j.shpsc.2005.03.008](#)
- (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford, UK: Clarendon Press.
- Darden, L. & Maul, N. (1977). Interfield theories. *Philosophy of Science*, 44 (1), 43-64.
- Darden, L. (1991). *Theory change in science: Strategies from mendelian genetics*. Oxford, UK: Oxford University Press.
- Fehr, C. (2004). Feminism and science: Mechanism without reductionism. *National Women's Studies Association Journal*, 16 (1), 136-156. [10.1353/nwsa.2004.0032](#)
- Floridi, L. (2008). The method of levels of abstraction. *Minds and Machines*, 18 (3), 303-329. [10.1007/s11023](#)
- Foster-Wallace, D. (2004). Consider the lobster. *Gourmet Magazine*, 64 (8)
- Gillett, C. (2002). The dimensions of realization: A critique of the standard view. *Analysis*, 62 (276), 316-323. [10.1111/1467-8284.00377](#)
- (2013). Constitution, and multiple constitution, in the sciences: Using the neuron to construct a starting framework. *Minds and Machines*, 23 (3), 209-337. [10.1007/s11023-013-9311-9](#)
- Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44 (1), 49-71. [10.1007/BF00172853](#)
- Hafting, T., Fyhn, M., Molden, S., Moser, M. & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436 (7052), 801-806. [10.1038/nature03721](#)
- Harinen, T. (forthcoming). *Causal and constitutive explanation*.
- Haug, M. C. (2010). *Realization, determination, and mechanisms*. *Philosophical Studies* 150 (3), 313-330.
- Hempel, C. G. (1965). *Aspects of scientific explanation*. Princeton, NJ: Princeton University Press.
- Hitchcock, C. (2003). Of Humean bondage. *The British Journal for the Philosophy of Science*, 54 (1), 1-25. [10.1093/bjps/54.1.1](#)
- Kauffman, S. A. (1971). Articulation of parts explanation in biology and the rational search for them. In R. C. Buck & R. S. Cohen (Eds.) *PSA 1970* (pp. 257-272). Dordrecht, NL: Reidel.
- Kim, J. (1998). *Mind in a physical world*. Cambridge, MA: MIT Press.
- (2000). Making sense of downward causation. In P. B. Andersen, C. Emmeche, N. O. Finnemann & P. Voetmann Christiansen (Eds.) *Downward causation. Minds, bodies and matter* (pp. 305-321). Aarhus, DK: Aarhus University Press.
- Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (Eds.) *Scientific explanation* (pp. 410-505). Minneapolis, MN: University of Minnesota Press.
- Lewis, D. (1991). *Parts of classes*. Oxford, UK: Blackwell.
- (2000). Causation as influence. *Journal of Philosophy*, 97 (4), 182-197.
- Machamer, P. K. & Sullivan, J. (2001). Leveling reduction. *University of Pittsburgh Philosophy of Science Archive*
- Machamer, P. K., Darden, L. & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 57 (1), 1-25. [10.1017/CBO9780511498442.003](#)
- Mackie, J. L. (1980). *The cement of the universe: A study of causation*. London, UK: Clarendon Library of Logic and Philosophy.
- Marcus, E. A. (2006). Events, sortals, and the mind-body problem. *Synthese*, 150 (1), 99-129. [10.1007/s11229-004-6258-7](#)
- Marr, D. (1982). *Vision*. San Francisco, CA: Freeman Press.
- Melnyk, A. (2010). Comments on Sydney Shoemaker's "Physical realization". *Philosophical Studies*, 148 (1), 113-123. [10.1007/s11098-010-9500-9](#)
- Moser, E. I., Kropff, E. & Moser, M. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annual Review of Neuroscience*, 31, 69-89. [10.1146/annurev.neuro.31.061307.090723](#)
- Oppenheim, P. & Putnam, H. (1958). Unity of science as a working hypothesis. In H. Feigl, M. Scriven & G. Maxwell (Eds.) *Concepts, theories, and the mind-body problem, Minnesota studies in the philosophy of science II* (pp. 3-36). Minneapolis, MN: University of Minnesota Press.
- Polger, T. (2004). *Natural minds*. Cambridge, MA: MIT Press.
- Rescher, N. & Oppenheimer, P. (1955). Logical analysis of gestalt concepts. *British Journal for the Philosophy of Science*, 6 (22), 89-106. [10.1093/bjps/VI.22.89](#)
- Romero, F. (forthcoming). Why there isn't interlevel causation in mechanisms. *Synthese*.
- Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.

- Samsonovich, A. & McNaughton, B. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *Journal of Neuroscience*, 17 (15), 5900-5920.
- Sanford, D. H. (1993). The problem of the many, many composition questions, and naive mereology. *Noûs*, 27 (2), 219-228. [10.2307/2215757](https://doi.org/10.2307/2215757)
- Schaffner, K. F. (1993). *Discovery and explanation in biology and medicine*. Chicago, IL: University of Chicago Press.
- Shagrir, O. & Bechtel, W. (forthcoming). Marr's computational-level theories and delineating phenomena. In D. Kaplan (Ed.) *Integrating psychology and neuroscience: Prospects and problems*. Oxford, UK: Oxford University Press.
- Shagrir, O. (2010). Computation, San Diego style. *Philosophy of Science*, 77 (5), 862-874. [10.1086/656553](https://doi.org/10.1086/656553)
- Shepherd, G. (1994). *Neurobiology*. London, UK: Oxford University Press.
- Simon, H. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.
- Thalos, M. (2013). *Without hierarchy: The scale freedom of the universe*. Oxford, UK: Oxford University Press.
- Van Gulick, R. (1993). Who's in charge here? And who's doing all the work? In J. Heil & A. Mele (Eds.) *Mental causation* (pp. 233-256). Oxford, UK: Oxford University Press.
- Varzi, A. (2014). Mereology. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/spr2014/entries/mereology/>.
- Wilson, R. A. & Craver, C. F. (2006). Realization. In P. Thagard (Ed.) *Elsevier handbook of the philosophy of psychology and cognitive science* (pp. 81-104). Dordrecht, NL: Elsevier.
- Wimsatt, W. (1974). Complexity and organization. In K. F. Schaffner & R. S. Cohen (Eds.) *PSA 1972* (pp. 67-86). Dordrecht, NL: Reidel.
- (1976). Reductionism, levels of organization, and the mind-body problem. In G. Globus, I. Savodnik & G. Maxwelll (Eds.) *Consciousness and the brain* (pp. 199-267). New York, NY: Plenum Press.
- (1997). Aggregativity: Reductive heuristics for finding emergence. *Philosophy of Science*, 64 (4), 372-384. [10.1086/392615](https://doi.org/10.1086/392615)

Mechanistic Emergence: Different Properties, Different Levels, Same Thing!

A Commentary on Carl F. Craver

[Denis C. Martin](#)

In this commentary I will briefly sketch the notion of “levels of mechanisms” as presented by Carl Craver and propose that we extend it to a more general notion of “level” that ensures wider applicability. The account of levels I develop is essentially based on an account of “properties”, claiming dependence of instantiation on a certain epistemic context. The main goal is then to reconcile Craver’s notion of “mechanisms” with “emergence” resulting in a contemporary account of “mechanistic emergence” implementing the developed concept of a level. Such an account could provide explanatory potential for and elucidate on seemingly mysterious higher-level properties and their ontology.

Keywords

Causation | Descriptive pluralism | Dispositional essentialism | Dualism | Emergence | Epistemic context | Level | Mechanisms | Mechanistic emergence | Mechanistic explanation | New essentialism | Novelty | Ontological novelty | Part-whole relationship | Properties | Property instantiation | Realization | Unexplainability | Unpredictability

1 Introduction

Are mind and brain on the same level? Mental properties and biological properties are so different that some kind of dualism is still an attractive position for many people. Intuitively, mental phenomena are often assumed to be on some kind of higher level than physical phenomena. For example, in order to accurately describe what it means to have compassion for another living being, most people would probably

agree with the popular expression that this simply *cannot amount to nothing but* the description of the underlying neurophysiological activity or behaviour *related* to that compassion—that presenting the neurophysiological activity alone does not fully capture all properties of being in a state of compassion. Instead, especially in everyday life, we might rather refer to the phenomenological properties of compassion,

Commentator

[Denis C. Martin](#)

denis.martin@hu-berlin.de

Humboldt Universität zu Berlin
Berlin, Germany

Target Author

[Carl F. Craver](#)

ccraver@artsci.wustl.edu

Washington University
St. Louis, MO, U.S.A.

Editors

[Thomas Metzinger](#)

metzinger@uni-mainz.de

Johannes Gutenberg-Universität
Mainz, Germany

[Jennifer M. Windt](#)

jennifer.windt@monash.edu

Monash University
Melbourne, Australia

the properties we draw on to identify that we are in a state of compassion at a given time. These properties seem to have a special value for us. In a way, they seem to be much richer than those of “cold” science. But what, exactly, does it mean to say that mental and physical phenomena are not on the same level? If I were to ask what compassion is, most people would probably agree that it is somehow realized by their body, just as an elaboration of this fact does not suffice for a complete description of compassion, implying that there must be something *more* than that, on a *higher* level. This, at least for a philosopher, inevitably leads to the question of what those levels actually *are*. What does “level” refer to? To what extent do levels *exist* in the world at all? These questions become even more pressing when we make ourselves aware of the extent to which the sciences use the concept of “level”. Whole disciplines, such as psychology and neuroscience, are distinguished as operating on different levels with different theories aiming at specific target phenomena. Levels also play a role within disciplines. In neuroscience, for example, it is quite common to distinguish between lower-level brain functions as realized in the brain stem or primary sensory areas as opposed to higher-level functions like decision-making or emotion regulation that are attributed primarily to the frontal lobe. Likewise, the distinction of processes and functions as operating “bottom-up” or “top-down” is quite prevalent.

There is a general strategy in science that has proven to be effective for explaining a certain phenomenon: decomposition. The reason for that is as follows: to fully explain a phenomenon, it does not suffice for us to be able to elaborately describe it or list certain correlations with singular components or other phenomena. Rather, we need to know in detail how the phenomenon comes into existence, based on how exactly it is realized: which components underlying the phenomenon are doing what, where, when, and how in order to make the phenomenon emerge. These requirements are captured excellently by Carl Craver’s (2007; Craver & Darden 2013, p. 15) famous definition of a mechanistic explanation:

mechadef/mechanistic explanation =_{Df} [m]echanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions. (Craver & Darden 2013, p. 15)

But in what sense are the mechanistic components of a phenomenon on a lower level than the whole phenomenon? This is the question Craver answers in his article “Levels”, in this collection.

In what follows, I shall first point out I find most important about Craver’s account of levels of mechanisms and where I see some difficulties in his account. I shall then propose an alternative way of defining levels by emphasizing the notion of “properties”. The idea here is that levels are a direct result of property instantiations and thereby constitute “property-dependent epistemic dimensions”. By focusing on properties in general, and not only on properties of mechanisms, I hope to show that an account of levels does not have to be as restricted as Craver proposes. I shall also argue that levels of mechanisms and levels of emergence do not have to be conceived as necessarily distinct, but can rather be combined quite well into a productive account of mechanistic emergence. Expanding Craver’s account of levels this way provides not only a notion of levels with wider applicability but also builds on his account of mechanisms as operating on different levels, which instils explanatory potential into a contemporary account of emergence. This still secures the application of “levels” in science, but at the same time makes transparent how the epistemic contexts of science are property- and level-generative. The ultimate goal of this approach is, of course, to elucidate how one and the same material system may show significantly different properties to an extent that elicits serious confusions about matters of identity.

2 Levels of mechanisms

First of all, the approach Craver takes for defining levels is notable in many ways. What I find especially important, however, is that he develops his definition in an interdisciplinary frame-

work, paying close attention to compatibility with or even application in neuroscience. So what are levels as used in such a scientific context? In his article of this volume, Craver extends his original definition of mechanisms a little to accommodate for the existence of lower-level mechanisms that take part in the realization of higher-level mechanisms.

I use the term ‘mechanism’ permissively to describe non-aggregative compositional systems in which the parts interact and collectively realize the behavior or property of the whole. Mechanisms are by definition more than the sums of their parts: they have properties their parts do not have, and they engage in activities that their parts cannot accomplish on their own. (Craver [this collection](#), p. 16)

Mechanisms as construed here are entities and activities organized in non-aggregative compositional systems, such that they are productive of regular changes from start or set-up to finish or termination conditions, and the properties of the whole mechanism are produced collectively through the interaction of its component mechanisms. This establishes the basis for Craver’s introduction of levels of mechanisms.

Craver’s three defining questions—namely about the *relata*, the relations, and their placement—constitute a valuable contribution to the conceptual clarification of the term “level” helping to capture the criteria for the correct usage of the term. This already hints at the descriptive pluralism Craver promotes, meaning that there is a set of equally correct ways to use the term depending on the respective answers to these three key questions. In the case of mechanisms, levels are best individuated, according to Craver, in terms of a part–whole relationship between the property “ ψ -ing” of S, given that S is a mechanism as a whole, and the property “ ϕ -ing” of X, given that X is a mechanism component that is a constitutively relevant spatiotemporal part of S.

Summing up Craver’s position on levels and levels of mechanisms as I understand it, they:

- are *metaphors* with multiple distinct conditions of correct usage dependent on the *relata*, relations, and their placement and
- refer to *part–whole relationships*.
- Levels of mechanisms have *properties* of mechanisms and properties of their parts *as their relata* (as opposed to levels of size, which have objects as their *relata*),
- are always *non-aggregative* (though aggregative levels do also exist),
- are *not monolithic*, but constitute a *local organizational part–whole hierarchy*, while
- the part–whole relationship must satisfy the *constitutive relevance condition*;¹ **while**
- there is *no causal relation* between them,
- they bear *explanatory potential*,
- and, finally, the *placement* of entities on levels of mechanisms *is weak* in the sense that for all entities that are not related as part and whole it can be said that they are on the same level.

What might the difficulties with this account of levels be? As a minor point, first, there might be some implications of using the concept “metaphor” in connection with levels. By definition, a “metaphor” identifies two things—a primary and secondary subject—with one another, such that one of the two can be captured in description more powerfully (Hills 2010). What could the primary and secondary subject in a level-metaphor be? The primary subject would probably be a level in the sense of a level of mechanisms, the secondary subject could maybe be a plane, a stage, or a degree. But how would that help elucidate what the primary subject levels actually are? As far as I can see, using “level” as a metaphor would more effectively *describe* what a level is, but not actually *define* it and, thereby, simply capture what our intuition about levels is in the first place—namely that it is somewhat analogous to a level in the secondary subject sense. Also, it seems that conceiving of “levels” as a metaphor would

¹ “Constitutive relevance”: “[...] all the lower-level properties, activities, and organizational features of the parts are relevant to—contribute to—the property or activity of the whole” (Craver [this collection](#), p. 15).

already somewhat negatively answer the question of what levels actually are, as solely existent as a figure of speech—an analogy that could be eliminated without any ramifications. Craver’s descriptive pluralist approach is formulated specifically to counteract elimination of levels and thereby to sustain their application in science. However, descriptive pluralism obviously does still act on the assumption that levels are metaphors, and only describes conditions that fit their application better than others. Therefore, this approach does not seem to be particularly helpful for our intended and certainly desirable goal.

Another issue I would like to raise is that in Craver’s account of levels as presented here, the key defining feature of levels seems to be a relation condition² between certain entities. This already becomes apparent with the three questions mentioned above, which aim to help us adequately describe specific instances of levels. Levels of mechanisms in particular are specified as a part–whole relationship between properties of mechanisms that are located on different levels. But does this really capture what *levels* actually are? Or does it rather render the relation condition *between* levels more precise, instead? While helpful to set the criteria for conditions under which the term “level” is correctly employed, and highlighting distinguishing features of different levels, saying that levels are essentially relations between sets of entities is at best an indirect or descriptive definition of levels and does not seem sufficient for a complete definition. It leaves open how levels come into existence, what their ontological status is, and why we posit certain entities on the same level in the first place, that is, what the commonalities of entities are that lead them to be on the same level.

From the key aspect of a part–whole relationship in Craver’s account stems the notion that levels are local and non-monolithic. This means that only those entities that are involved in such part–whole relationships can intelligibly be said to be on different levels. What are our theoretical options for conceptually covering all

other entities? Since they don’t fulfil the part–whole criterion they cannot be on distinct levels and, therefore, in a sense they are all on the same level. However, according to the definition of levels at hand, to be on any level means that there are other levels, which are distinguished from the first level by the part–whole relationship of the entities involved. Since the entities under consideration do not exhibit this kind of relationship, they are on no level at all. In general, this seems like a reasonable option. But let’s consider the case that entities that are not in such a part–whole relationship do, nevertheless, share some features—for example a hedge and a fence both one meter high. In accordance with levels not being monolithic and the previous considerations, these would not be on the same level, but on no level at all. What exactly is wrong, however, with saying that two entities sharing the feature of being one meter high but which are not related as part and whole are on the same level? One could, of course, simply invoke the account of levels of mechanisms and argue that it is designed such as not to warrant such a level. But does this limitation really procure us a better understanding of “level”, or could it rather be too restrictive for that purpose? Its consequence, at any rate, is a very strong focus on the vertical dimension, namely the relation between levels, whereas the horizontal dimension, that is, entities related qua being on the same level, is somewhat neglected. So, let us ask, what are the criteria for two entities being on the same level? It is exactly this relation between entities on the *same* level that the concept is primarily supposed to capture, and yet which seems to be underspecified by the definition provided. And how similar do two part–whole relationship units have to be in order for it to be correct to say that their respective wholes and parts are on the same level? Or does it even follow that two things that are not part of one and the same part–whole relationship cannot even be on the same level at all?

As a third point, finally, there remains the issue of the extent to which levels of mechanisms are similar or distinct from levels of realization and emergence, respectively. All three kinds of levels share that an application of the

² “Relation condition”: “the componency relationship between things at higher and lower levels” (Craver [this collection](#), p. 19).

concept of inter-level causation is not feasible in their case, since they do not fit classical assumptions about causation such as non-synchronicity of cause and effect—a very substantial point Craver emphasizes in the target article. But what are the differences between these kinds of levels?

According to Craver, levels of mechanisms and levels of realization seem to differ in that the former exhibit a relationship between wholes and parts, whereas the latter exhibit a relationship between wholes and sets of realizers. But this distinction seems rather frail. How are the parts involved in levels of mechanisms different from the set of realizers involved in levels of realization, such as to warrant this distinction? At least in levels of mechanisms, as Craver envisages them, the parts are several mechanisms that together form the whole, which is comprised of all the particular “part-mechanisms”. If mechanisms in general realize certain phenomena, this suggests that all “part-mechanisms” are also realizers in the same way. Now, if the “part-mechanisms” on the “part-levels” are all realizers, it is reasonable to say that the “whole-mechanism” on the “whole-level” is realized by the organized coaction of its parts and that the “part-levels” are the realizers of the “whole-level”. Thus, the distinction between levels of mechanisms and levels of realizers conflates.

The difference between levels of mechanisms and levels of emergence, on the other hand, is based on the unpredictability, unexplainability, and metaphysical novelty of higher-level properties, as opposed to lower-level ones. Craver’s point here is that levels of mechanisms, while they can be unpredictable, do not have to be so necessarily, that they are always explanatory, and the novelty of higher-level properties is a trivial fact. But why think that the opposite must hold in the case of emergence? Of course, “spooky emergence”, “[...] the existence of higher-level properties that have no explanation in terms of the parts, activities, and organizational features of the system in the relevant conditions” (Craver [this collection](#), p. 21), is spooky by definition—that much is clear. Also, admittedly, the way emergence was construed

historically by the British Emergentists perfectly fits this view and deliberately opposes mechanisms as it can be found, for example, as per Broad (1925). However, why should we prematurely accept this view of emergence as given and eliminate any possibility of further development towards a notion of emergence that is perfectly commensurable with modern science? In fact, I think the formidable way in which Craver develops his account of levels of mechanisms is perfectly suited to facilitate development in this direction. So what could a definition of emergence be, and how can it be united with mechanisms? The following definition of emergence by Evan Thompson (2007) already seems compatible with Craver’s framing of the way properties of higher-level mechanisms are constituted by properties of lower-level ones.

A network, N , of interrelated components exhibits an emergent process, E , with emergent properties, P , if and only if:

- (1) E is a global process that instantiates P , and arises from the coupling of N ’s components and the nonlinear dynamics, D , of their local interactions.
- (2) E and P have a global-to-local (“downward”) determinative influence on the dynamics D of the components of N .

And possibly:

- (3) E and P are not exhaustively determined by the intrinsic properties of the components of N , that is, they exhibit “relational holism.” (p. 418)

This definition is compatible with Craver’s characterization of levels of mechanisms in the following respects: properties of higher-level mechanisms, global emergent properties, are realized by properties of lower level mechanisms; and there is a part–whole relationship between those relata, as well as a non-causal influence between the levels. What Thompson’s definition additionally contributes is a point about predictability. For many phenomena in

nature, the interactions of lower-level components are so complex that they can only be described by non-linear dynamics. A precise predictability of the higher-level phenomena might not always be possible at present due to there being too many factors involved in the underlying processes—it simply exceeds current computational tractability. Craver acknowledges this point—so it can also be said to be consistent with his account—but worries that this might have ontological ramifications: “[i]f that explanatory relationship is severed, then the sense in which emergent properties are at a ‘higher-level’ must be altogether different than the compositional notion of levels in levels of mechanisms” (Craver [this collection](#), p. 21). More precisely, he suspects that in emergence ontological novelty arises through the epistemological limitations just mentioned; otherwise ontological novelty would simply be a banal fact already expressed by his account.

There are, however, several problems with this view: first, one can make a distinction between the epistemological (e.g., “predictability”) and ontological (e.g., “novelty”) dimension of emergence (O’Conner & Wong 2009). There is, in principle, no reason to assume that the ontological dimension is dependent on the epistemological; rather, they seem to be fully dissociable.

Second, his criticism backfires with regard to the banality of the properties that higher-level mechanisms exhibit in his own account of levels. For there to be “higher” levels of mechanisms, these mechanisms must show *new properties*, that is, in order for them to qualifiedly be on that level. Hence, his account cannot go without ontological novelty of some kind. To now say that this ontological novelty would be only a banal fact undermines his very own striving for mechanistic explanation, which certainly is not banal. In fact, it is still interesting how “higher-level” properties come into existence, what it means to say that they are *new*, and how the concept “level” might be connected to this. A successful reconciliation of mechanisms and emergence in the form of mechanistic emergence could provide a solution to this problem.

Third, the dissociability of the epistemological and ontological dimension of emergence does not contradict the possibility of their mere coexistence. Once we dismiss the idea that ontological novelty follows from epistemological intractability, overcoming the restraints of historical accounts of emergence, the fact that the coming into existence of new properties on a higher level is not tractable at the moment does not mean that it is not so *in principle*. The reason why we call this coming-into-existence “emergence”, as might be conceived by a revised account, is not based on the fact that it *is* epistemically intractable *in principle*, but rather that it shows novel properties on a higher level that *appear* to be epistemically intractable *in principle*, while they might at some point be very well explained in a mechanistic framework combined with a proper theory of property instantiation.

Thompson’s definition leaves room for local components to be part of a mechanism. A mechanistic explanation of the emergent phenomenon, it seems, would not be incompatible with an account of emergence, but rather contribute to its explanation by elaborating on how the organization of the parts is essential for emergent properties to arise. As for the ontological novelty of higher-level phenomena, this is certainly a crucial point in emergence: there are new properties coming into existence on the higher levels that are somehow realized by processes of components on the lower level, which in isolation do not show the same properties as the whole. For this to happen, however, contrary to what Thompson’s definition implies, the underlying interactions of the components or the emergent properties themselves must not necessarily be unpredictable in principle. But I anticipate an objection: this form of emergence would again only be very weak or banal, but not ontologically new. As already mentioned above, unpredictability does not have a bearing on ontological novelty and is therefore not crucial for emergence. What levels of mechanisms and emergent levels share is that on higher levels there are *new properties*, which means that there is a notable ontological difference. How extensive such an ontological difference

must be in order not to be banal remains a matter of debate. Still, in the case of mechanisms, as well as in the case of emergence, it is very likely that there is a significant phenomenon making up the higher level—otherwise it probably would not be of such interest for enquiry as it clearly is.

As a result of these thoughts, in what follows I will try to reconcile levels of mechanisms and levels of emergence as two interconnected forms of realization. This alone, of course, does not solve the problem of what the levels involved actually *are*. So in fact there are two problems to be solved for an account of mechanistic emergence: (a.) what it means exactly for a higher-level phenomenon to exhibit *new* properties, and (b.) what exactly constitutes a *level*. As a route to a possible solution, in the next section I will sketch a definition of properties that can be implemented in a definition of mechanistic emergence and that at the same time provides a positive account of levels.

3 Level-carving properties in mechanistic emergence

In this section, I propose an alternative account of levels that is fully compatible with the mechanistic framework and the way in which levels of mechanisms are construed by Craver, but which at the same time has a wider scope of application. Since this account will rely on properties as the crucial defining criterion, I shall first sketch a working definition of the concept of “properties”. In a second step, this definition of “property” will conceptually ground the alternative account of levels. Finally, I will implement both the definition of properties and that of levels into a formulation of mechanistic emergence.

What might be a working definition for the term “properties”? Inspired by [Brian Ellis’ \(2002\)](#) “new essentialism”³ and [Alexander Bird’s \(2007\)](#) “dispositional essentialism”,⁴ I

propose the following view: what exist in the world are entities with individual dispositional profiles. An example that [Ellis \(2002\)](#) gives is the dispositions of particles to attract or repel each other. These essential dispositions, individuating the particles as that which they are, make it possible for us to formulate laws of nature. Many of these dispositions are the result of structural and organizational combinations of matter with different dispositions, e.g., ions have the disposition to form ionic crystals, which by means of the resulting structural characteristics in turn have other specific physico-chemical dispositions. Those essential dispositions alone, however, are not properties yet. That is because dispositions exist outside of an epistemic context. The property of being one meter high, for example, is dependent on the disposition of an object to exactly fit the measurement revealing it to be the height of one meter. Without the measurement, however, the property is not instantiated—only the disposition of its instantiation exists inherently in the structure of the object. Thus, according to my theory, properties are instantiated through the interaction of the essential dispositions of matter and an epistemic system. Of course, now you will ask what this epistemic system could possibly be. Admittedly, this aspect of the definition is in a particularly embryonic stage and requires further research. As a general characterization, epistemic systems are structured such that they feature sensors or gauges that capture specific dispositions of entities and provide characteristic values as an output. Human and non-human animals, as well as physical devices of measurement, are epistemic systems in this sense. Let us note the following as a working definition of “epistemic system”:

Epistemic system =_{Df} (ES) Epistemic systems are organized (a.) such as they feature sensors or gauges that pick up a specific disposition exhibited by an entity and (b.) such as there is a transformation of that signal into a particular value characteristic of the system’s organization.

³ “[...] things must behave in the sorts of ways they do not because the laws of nature require them to, but rather because this is how they are intrinsically disposed to behave” ([Ellis 2002](#), pp. 3–4).

⁴ “[...] the claim that fundamental natural properties are essentially dispositional. [...] *x* is disposed to manifest *M* in response to stimulus *S* iff were *x* to undergo *S* *x* would yield manifestation *M*” ([Bird 2007](#), p. 24).

With this definition at hand we are able to formulate a working definition of properties:

Properties =_{Df} (P) Properties are instantiated through epistemic processes, which are constituted by interactions between epistemic systems and complementary dispositional profiles of entities.

Let us now turn to how levels depend on properties. The idea here is that levels are established by the epistemic systems in use that instantiate the properties which belong to the respective level. Measuring ion conductance at an axon with electrodes, for example, establishes properties on a cytological level; whereas measuring reaction times of participants in a behavioural experiment establishes properties on a psychological level. The way in which different epistemic systems — e.g., a functional magnetic resonance imaging (f) scanner and a blood test — applied on the same entity — a human — establish different properties — a local decrease in de-oxygenated hemoglobin versus, for instance, cortisol levels in the bloodstream — on different levels — the level of brain activation versus the level of endocrine activation — shows that levels and properties are intimately connected. Coming back to our example from the beginning, different properties of the mental state compassion are instantiated in several ways: (a.) through a person as an epistemic system directed towards a myriad of dispositional interactions of her own body, which can then (b.) be picked up as values of standardized questionnaires probing those experiences while, finally, (c.) some properties of the underlying mechanism of compassion are instantiated by means of fMRI (cf. Klimecki et al. 2014). What becomes strikingly apparent in this example is that each of these different ways of property instantiation yields properties we would intuitively base on very different levels. Experience of compassion seems to have a very different quality and complexity to the more abstract numerical values of questionnaires or activation patterns visualized by fMRI. Thus, we can say that the specific way a certain property is instantiated already establishes a corresponding level. This way of defin-

ing levels offers a broader range of application than the levels of mechanisms account, since it is not restricted to properties of mechanisms but rather bears on properties in general. After these considerations, we are now in a position to put down the following as a working definition of levels:

Levels =_{Df} (L) Levels are sets of properties established with respect to their instantiation through the same or a similar kind of epistemic system, which targets the same or a similar dispositional profile as compared to different epistemic systems targeting different dispositional profiles.

Considering the identity criteria of epistemic systems, an epistemic system identical to itself might be involved in the instantiation of two properties, that are on the same level qua being picked up by the said epistemic system—for example, a ruler instantiating the length of 1 meter and 20 centimetres. Two epistemic systems are similar if they pick up exactly the same kind of dispositions and exhibit a similar dimension of output value. For example, two rulers picking up the dispositions of a set of tables to instantiate the height of 1 meter and 59.1 inches (1.5 m) or two humans seeing the color blue. Note that the properties might be different in these cases but they are still on the same level, since they are instantiated by the same or similar epistemic system. A ruler and an infrared detector, for instance, are neither identical nor similar epistemic systems, since they differ in the kind of dispositions they pick up and have different dimensions of output values.

Having provided the two missing definitions for an account of emergence, let us now consider how these can be connected to mechanisms and implemented into a new account of mechanistic emergence. What could properties of mechanisms be? According to the definition of mechanisms given in the introduction, we have to identify entities and activities belonging to the mechanism, as well as starting and termination conditions. These might all be established by property-instantiating epistemic sys-

tems. We decompose, measure, and intervene with the phenomena or their realizing components to establish temporal, functional, or organizational properties. Properties of mechanisms on higher levels are instantiated differently to properties of mechanisms on lower levels. We can use one epistemic system to track particular sequences within a mechanism on a specific level in order to be able to recognize stepwise changes as they unfold in temporal order. For properties on higher levels, however, we have to change the kind of epistemic systems involved. We are dealing with different properties on a different level, and we cannot capture the same causal chain as in the single lower-level mechanism. Instead, we capture a synchronously emergent property on a higher level. A formulation of such an account of mechanistic emergence that incorporates the above definitions of “properties” and “levels” could be constructed as follows:

Mechanistic emergence $=_{\text{Df}}$ (ME) Mechanistic emergence is a special form of property instantiation, in which the novelty of the properties is established by a change in epistemic systems involved in their instantiation and through which they span a higher level compared to the level of the components, which realize the higher-level properties by means of their mechanistic organization and process dynamics, thereby changing the overall dispositional profile of the whole while, at the same time, being constrained with respect to their individual dispositions in virtue of the synchronous, non-causal constituency relation.

What is new in this account of emergence is that it acknowledges how even emergent properties ultimately arise out of perfectly explainable mechanistic processes. Unpredictability or unexplainability are no longer the defining characteristics of emergent properties themselves, but only characteristics of the *epistemic context* involved in their instantiation. However, these emergent properties are still *novel* in the sense that they are non-causal, non-aggregative

products of mechanisms that come into existence only on a higher level, established through the kind of property instantiation that none of the components show in isolation.

4 Conclusion

In this commentary my aim has been to point out (i.) that defining levels as crucially dependent on properties has a wider and more flexible range of application than using part–whole relationships as the defining criterion; and to put forward (ii.) that levels of mechanisms and levels of emergence can be reconciled into an account of mechanistic emergence in which the property-dependent definition of levels finds application.

My argument was (a.) that descriptive pluralism, by conceiving of levels essentially as metaphors, cannot yield sufficient conceptual clarification concerning the term “level”—namely, what levels actually are and how they exist and, even undermines the goal of preserving the use of the concept in science. Further, I highlighted (b.) that in “levels of mechanisms”, as presented by Carl Craver, the core criterion for a definition of “level” is a part–whole relationship in conjunction with a constitutive relevance constraint, and that this focuses solely on the vertical dimension existing between levels, and completely omits the more important horizontal dimension of the conditions that must apply for a set of entities to be on the same level. As such, the concept is only very weakly and indirectly characterized, offering little toward its clarification and broader application. Finally, I showed (c.) that ontological novelty is not dependent on epistemic intractability, and that the ontological novelty of properties on higher levels of mechanisms is not a banal fact either in levels of mechanisms or in levels of emergence. What emergence expresses at its core is that new properties are coming into existence and that they are so strikingly novel that they might not be predictable at the moment—or seem to not be so, even in principle. They are novel to such a degree that their instantiation coinstantaneously establishes a wholly new level.

As positive proposals for an alternative view, I defined (d.) “properties” as instantiated by epistemic systems capturing specific dispositional profiles of entities, and (e.) “levels” as sets of such properties instantiated by the same or a similar epistemic system, as compared to those properties instantiated by another epistemic system. Furthermore, I (f.) provided a definition of “mechanistic emergence” implementing the core idea of emergence as aforementioned, together with the proposed definitions of “levels” and “properties”.

Concerning future directions for research, it seems most pressing to further develop the notion of an “epistemic system”. Moreover, the notion of “levels” needs to be further refined with regard to how much epistemic systems can or must differ in order for there to be a new level. Ultimately, of course, it will be intriguing to see whether the developed definitions hold in the light of practical implementations in scientific contexts.

To finish, let us come back to the initial question of whether mind and brain, or more precisely mental processes and neurobiological processes, are on the same level, which we are now in a better position to answer. So far as we acknowledge that mental properties and properties of the brain are different properties, and if we also consider how I defined levels above, we can conclude that mind and brain are in fact on different levels in this sense. But are we thus slipping back into dualism? Absolutely not, since the definition of properties developed above makes clear how it comes about that there can be different properties of one and the same thing: it is dependent on the kind of epistemic system in use to capture specific dispositions or, in short, on the *epistemic context*. Taking up our example of compassion once again, it is now obvious why the phenomenon did not seem to be fully captured only by reference to, for instance, physiological properties. Of course it makes sense to investigate the physiological realization of compassion, to measure autonomic parameters, conduct blood tests, or undertake fMRI scans, but it is equally important to conduct behavioural experiments or even interviews with participants to target the phe-

nomenological experience that encompasses compassion (cf. Singer & Bolz 2013). This only means that we are doing research on all the different properties of the phenomenon of compassion. We are doing research in different disciplines with different methods, on different levels, and we are capturing different properties of one and the same thing—so let’s work together to incrementally integrate those epistemic contexts and get the complete picture.

Acknowledgements

I wish to express my gratitude to two anonymous reviewers, and in particular the editors, for their extensive comments on an earlier version of this commentary, which contributed immensely to its reaching the present form.

References

- Bird, A. (2007). *Nature's metaphysics: Laws and properties*. Oxford, UK: Clarendon Press.
- Broad, C. D. (1925). *The mind and its place in nature*. London, UK: Routledge & Kegan Paul.
- Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford, UK: Clarendon Press.
- (2015). Levels. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- Craver, C. F. & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. Chicago, IL: University of Chicago Press.
- Ellis, B. (2002). *The philosophy of nature: A guide to the new essentialism*. Chesham, UK: Acumen.
- Hills, D. (2010). Metaphor. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*.
<http://plato.stanford.edu/archives/win2012/entries/metaphor/>.
- Klimecki, O. M., Leiberg, S., Ricard, M. & Singer, T. (2014). Differential pattern of functional brain plasticity after compassion and empathy training. *Social cognitive and affective neuroscience*, 9 (6), 873-879.
[10.1093/scan/nst060](https://doi.org/10.1093/scan/nst060)
- O'Conner, T. & Wong, H. Y. (2009). Emergent properties. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*.
<http://plato.stanford.edu/archives/spr2012/entries/properties-emergent/>.
- Singer, T. & Bolz, M. (Eds.) (2013). *Compassion: Bridging practice and science*. Leipzig, GER: Max Plack Institute for Human Cognitive and Brain Sciences.
- Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge, MA: Belknap Press of Harvard University Press.

Mechanisms and Emergence

A Reply to Denis C. Martin

Carl F. Craver

I respond to some of the major issues in Martin's commentary, particularly (i) his insistence on a robust notion of being "at" a level, and (ii) his desire for mechanistic emergence to explain the genuine ontological novelty of higher level phenomena.

Keywords

Emergence | Levels | Mechanisms

Author

[Carl F. Craver](#)

ccraver@artsci.wustl.edu

Washington University
St. Louis, MO, USA

Commentator

[Denis C. Martin](#)

denis.martin@hu-berlin.de

Humboldt Universität zu Berlin
Berlin, Germany

Editors

[Thomas Metzinger](#)

metzinger@uni-mainz.de

Johannes Gutenberg-Universität
Mainz, Germany

[Jennifer M. Windt](#)

jennifer.windt@monash.edu

Monash University
Melbourne, Australia

1 Introduction

The goal of my target essay is to articulate and recommend a view about what one is and is not committed to when one asserts that the domain of neuroscience or some other special science spans multiple levels of organization. How useful one finds my specific recommendation—that they be understood as levels of mechanisms—will depend on the uses for which one deploys the level metaphor. Martin raises a number of questions about my unpacking of the metaphor that help to clarify what is at stake in this discussion.

2 The levels metaphor

First, a word about the idea that the term level is used as a metaphor. According to the Oxford English dictionary, the term "level" derives from terms describing the idea of being parallel to the ground. From there, the term could easily extend to the idea that a building can have multiple levels, or landings, from the ground up. Something like this spatial arrangement of stacked landings seems to be the basis for many applications of the level metaphor (in other words, landings are the "secondary subject" of the metaphor). Usage of the term "level" increased dramatically in the 20th Century, and surely some of this increase is explained by the

fecundity of the metaphor for ordering items in a set from the highest to the lowest (under some understandings of “high” and “low”). These are the specific topics (the primary subjects) to which the secondary subject (landings) is compared. In asserting that the term “level” has this metaphoric aspect, I do not intend to thereby disqualify the metaphor from also expressing a set of true or false commitments about the structure of the world. Indeed, I think that the mechanistic application of the metaphor more or less accurately describes the structure of at least many systems in the domains of the special sciences.

My point in being a pluralist about the levels metaphor is simply to emphasize that the metaphor of a horizontal landing is used to describe a wide variety of altogether distinct primary subjects. As a result, any understanding of how this level metaphor works, what it is committed to, and whether it works for a particular purpose, must begin with a clear understanding of how the metaphor is unpacked in the given application. (One cannot, to pick a perfectly clear example, equate the great chain of being and levels of mechanisms.) A second point was to emphasize that there are many legitimate applications of the metaphor, and that the utility of the metaphor might vary from one application to the next. I am not arguing for a single, correct way of understanding the level metaphor. However, I am arguing that one prominent use of the level metaphor in sciences such as neuroscience can be usefully unpacked as describing mechanistic levels.

3 Being “at” a level

Martin is particularly concerned with a) the idea that there should be a clear sense of what it is to be “at” a level, and b) the idea there is some significant sense in which being at a level represents a genuine ontological novelty.

The first of these concerns arises from the fact that the placement question cannot be (or has not been) given a satisfying answer within the mechanistic application of the metaphor. The mechanistic account of levels focuses, as Martin correctly notes, on what it means to be

at *different* levels of organization within a mechanism. Concerning what it means to be “at a level,” one can say only that two things are “at” the same level just in case they are not at different levels; if neither thing is a component entity/activity in the behavior of the other, then the two things are not at different mechanistic levels and, in this weak sense, they are at the same level. This provides a sort of answer to the placement question, but not one that will satisfy those, such as Martin, who hanker after some additional factor (such as size or similarity) that unites the items at a level and that explains why all such items are at that level. I take the failure to answer the placement question as one of the key revisionary consequences of thinking clearly about levels of mechanisms. It is a crucial guide to understanding what’s misleading about the monolithic conception of levels.

One apparent problem for this consequence (the absence of a satisfying answer to the placement question) was first raised by Lindley Darden (personal communication): X’s -ing might be a component in (and so at a lower level than) S’s ψ -ing, *and* yet both X’s -ing and S’s ψ -ing might be at the same level as (i.e., not at a different level than) some altogether distinct P’s β -ing. There is a failure of transitivity. This, it seems to me, is simply a consequence of the idea that the application of the level metaphor to levels of mechanisms breaks down when one is not talking about relations between parts and wholes. This is unproblematic; it is simply an alternative way of expressing the idea that being “at the same level” is of no additional metaphysical significance within the mechanistic application of the metaphor than simply not being at different levels. This does not prevent one, of course, from using some other ordering criterion for this expressive purpose; one might lump things together on the basis of their sizes or perhaps the instruments used to detect them. But one should be aware that at this point one has left the mechanistic application of the metaphor.

Another consequence of the mechanistic view is that one might equally correctly carve the boundaries of mechanisms in any number of

ways (see e.g., Craver 2004, 2009, 2012). If so, there will not be a uniquely correct answer to the question of how many levels a given mechanism has. Our decisions to privilege some grains in the decomposition of a mechanism as an appropriate place to locate a level depends on our techniques, our theoretical background assumptions, our characterization of the phenomenon, our representational tools, and perhaps certain features of human psychology. As Simon argued, these systems are only *nearly* decomposable; how a system is decomposed depends, for example, on the relative strength of intra- versus extra-system causal interactions that one takes to be appropriate for carving the system at its near-joints. This is another reason I am hesitant to pin too much on the idea of being “at” a level.

4 Ontological novelty and emergence

I am not hostile to the second idea (b) that levels of mechanisms exhibit a kind of ontological novelty, depending on how this novelty is understood and how one thinks that it is achieved. I suggest in the target article that a mechanism as a whole can do things that its parts (taken individually) cannot do. Lawnmowers mow grass; spark plugs do not. But I also claim that a mechanism as a whole cannot do things that its organized and interacting parts cannot do. This is because the behavior of a mechanism as a whole just is (or at least is ontologically intimate with) the organized interaction of its component parts. I don’t know how to make the notion of “ontological intimacy” precise here (though the behavior of a mechanism in context will surely supervene on the organized collection of interacting components in that context). In my hands, “ontological intimacy,” is meant to denote an exhaustive ontological grounding of the behavior of a mechanism as a whole in the organized interactions of its components in a given causal context, however that is properly to be unpacked. Everything has an ontic explanation in terms of the organized activities of the mechanism’s parts.

The term emergence strikes me as suspicious, and I hesitate to use it even in the con-

text of levels of mechanisms, precisely because it suggests a severing of this ontological intimacy, a slide from the banal fact that mechanisms behave as they do because they are organized arrangements of interacting parts to the ontologically esoteric thought that something comes into being with causal powers ontologically inexplicable in terms of the organized interactions of the parts. Such claims about ontological novelty are almost always accompanied by claims that emergent things have a “downward” causal influence. I have labored to drive a wedge between levels of mechanisms and such ideas.

This connects with Martin’s description of Thompson’s view. According to Martin, Thompson requires that the parts in a system must be coupled and dynamically interacting to produce emergent properties and, further, that the properties of the whole should act “downward” on the dynamics of the components (2007). Take these in turn.

The idea of levels of mechanisms, by itself, requires only that the parts be organized and interact with one another; it does not require that the components interact non-linearly. Mechanisms with components that interact non-linearly are a subset of mechanisms more generally. It should not make any difference with respect to the novelty of higher-level properties that the parts interact non-linearly. Such dynamical interactions might make the behavior of the whole harder to predict on the basis of our understanding of the parts; but this is an epistemic, not an ontological, observation irrelevant to the ontological question to which I am responding. To be ontologically novel is not merely to be surprising.

Perhaps systems with dynamically interacting components are harder to idealize into separable interacting components; again, this would appear to be an epistemic issue. Thompson, at least in Martin’s description, is drawing a more fine-grained distinction than I draw. This is perhaps a terminological matter. The pressing question between us is whether non-linearity (or perhaps some other form of complexity) makes any ontological difference to the kind of thing that “emerges” from the “organized interactions” of the parts. I’m inclined

to say no, but a full consideration of the issue would require a more detailed treatment than I can give it here.

Thompson also requires that the higher-level phenomenon should have “determinative” influence on things at lower-levels. “Determinative” surely cannot mean the existence of a universally quantified material conditional with the behavior of the mechanism as a whole as the antecedent and the organized interactions among the components as the consequent. Multiple realization precludes such an analysis. But “determinative” also cannot be understood in the causally productive sense in which something (such as charge or momentum) is passed from whole to part. And for reasons expressed in the target article, there are many widely shared assumptions about the independence of causes and effects that stand in the way of understanding such top-down relations in causal terms. This is why I suggest (as in [Craver & Bechtel 2007](#)) that the language of downward causation is misleading in the context of token mechanisms. We recommend that claims about downward causation are really shorthand ways of expressing an often complex web of constitutive (about relations between things at different levels of mechanisms and levels of realization) and causal claims (about things that are not at different levels from one another) about the system. However this is to be cashed out, the term “determinative” seems misleading.

5 Levels and techniques

Finally, I’ll offer a brief remark concerning Martin’s positive suggestion for reifying levels. Martin’s central idea—that we often pick out things as being on a level because they are detectable with the same or a similar kind of apparatus or with the same or a similar set of procedures—seems to me correct and important. Among the many possible ways of decomposing a spatio-temporal whole into component parts, some of these correspond to items that for one reason or another are readily detectable as such by one or more experimental apparatus. This way of putting things highlights how pragmatic factors, such as available apparatus, determine what we

will take to be an appropriate “landing” in a hierarchy of levels. And it raises useful questions about, for example, when two different apparatus, or two different tasks, or two different procedures in fact target the same items, or items at the same level. On a fine-grained characterization of our experimental instruments, no two experiments are the same. On a coarse grain, even superficially quite distinct experiments can be targeting the same phenomenon (consider, e.g., implicit bias or spatial memory). As noted above, it seems to me that many other epistemic, theoretical, and psychological factors enter into these decisions as well.

This good point, however, is obscured by an ontology in which properties apparently come into existence during acts of detection. Martin’s view is that objects have disposition profiles, and these profiles are turned into properties when they are measured. It is unclear to me, however, why one would not want to say that properties are there to be detected all along, or perhaps that properties just are the dispositional profiles of things. Martin doesn’t do much to motivate this experimental idealism about properties, but the thought seems hard to motivate, at least for macroscopic phenomena. It is an apparent consequence of Martin’s account that levels don’t exist until they are detected. And it is an apparent consequence of his view that new levels come into existence when we develop new instruments to detect them. And, to reiterate the thought in the last paragraph, we will have to wrestle with the question of when two techniques detect the same thing or different things. Martin’s positive proposal makes this question much more pressing, given that, for Martin’s view, the structure of the world—specifically, the distribution of properties in space and time—apparently hangs in the balance. But if we abandon the idea that the placement question must have a uniquely correct answer, these questions are less pressing for thinking about ontology; nonetheless, questions about how we coordinate different experimental tasks, protocols, and procedures are at the heart of the epistemological challenge faced by any experimentalist (see [Sullivan 2009, 2010](#)).

References

- Craver, C. F. (2004). Dissociable realization and kind splitting. *Philosophy of Science*, 71 (5), 960-971.
- (2009). Mechanisms and natural kinds. *Philosophical Psychology*, 22 (5), 575-594. [10.1080/09515080903238930](https://doi.org/10.1080/09515080903238930)
- (2012). Functions and mechanisms: A perspectivalist account. In P. Huneman (Ed.) *Functions* (pp. 133-158). Berlin, GER: Springer.
- Craver, C. F. & Bechtel, W. M. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, 22 (4), 547-563. [10.1007/s10539-006-9028-8](https://doi.org/10.1007/s10539-006-9028-8)
- Sullivan, J. (2009). The multiplicity of experimental protocols: A challenge to reductionist and non-reductionist models of the unity of neuroscience. *Synthese*, 167 (3), 511-539. [10.1007/s11229-008-9389-4](https://doi.org/10.1007/s11229-008-9389-4)
- (2010). Reconsidering spatial memory and the Morris water maze. *Synthese*, 177 (2), 261-283. [10.1007/s11229-010-9849-5](https://doi.org/10.1007/s11229-010-9849-5)
- Thompson, E. (2007). *Mind and life: Biology, phenomenology, and the sciences of the mind*. Cambridge, MA: Belknap Press of Harvard University Press.