Levels: descriptive, explanatory, and ontological

Christian List*
LSE

March-April 2016

Abstract

Scientists and philosophers frequently speak about levels of description, levels of explanation, and ontological levels. This paper presents a framework for studying levels. I give a general definition of a *system of levels* and discuss several applications, some of which refer to descriptive or explanatory levels while others refer to ontological levels. I illustrate the usefulness of this framework by bringing it to bear on some familiar philosophical questions. Is there a hierarchy of levels, with a fundamental level at the bottom? And what does the answer to this question imply for physicalism, the thesis that everything supervenes on the physical? Are there emergent higher-level properties? Are higher-level descriptions reducible to lower-level ones? Can the relationship between normative and non-normative domains be viewed as one involving levels? And might a levelled framework shed light on the relationship between third-personal and first-personal phenomena?

1 Introduction

Scientists as well as philosophers frequently employ notions such as *levels of description*, *levels of explanation*, and *ontological levels*. Although it is widely held – though by no means universally accepted – that everything in the world is the product of fundamental physical processes, it is also widely recognized that, for many scientific purposes, the right

*This paper articulates a framework that has been implicit – to variable extents – in several of my previous works, but has never been spelt out in full. Relevant works include List and Menzies (2009, 2010), List and Pettit (2011), List and Spiekermann (2013), List (2014), List and Pivato (2015a,b), and Dietrich and List (2016, Section 8). I wish to record my intellectual debt to all of my co-authors of these works: the late Peter Menzies, Philip Pettit, Kai Spiekermann, Marcus Pivato, and Franz Dietrich. I am also grateful to Marcus Pivato for detailed comments on a draft of this paper, and to Richard Bradley, Elizabeth Coppock, Johannes Himmelreich, George Musser, Wlodek Rabinowicz, and Laura Valentini for helpful discussions. My work was supported by a Leverhulme Major Research Fellowship.

level of description or explanation is not the fundamental physical one, but a "higher" level, which abstracts away from microphysical details.[1] Chemistry, biology, geology, and meteorology would all get bogged down with an informational or computational overload if they tried to explain the phenomena in their domains by modelling the behaviour of every elementary particle, instead of invoking "higher-level" properties and regularities. For instance, it would be a hopeless task to try to understand a biological organism or an ecosystem at the level of the billions of elementary particles of which it is composed, rather than at the macroscopic level of its biological functioning.

Similarly, cognitive scientists tend to assume that human psychology is better understood at the level of the mind (the cognitive-psychological level) than at the level of the brain (the neuro-physiological level).[2] This parallels the observation that it is much easier to understand the workings of a word-processing package such as Microsoft Word at the software level than at the hardware level, where astronomical numbers of electrons flow through microchips.

Finally, for many social-scientific purposes, the right level of description is not the "micro"-level of individuals, but a social level, involving "macro"-variables.[3] Despite the popularity of methodological individualism – the view that social phenomena should be explained at the level of individuals – macro-economists and political scientists would have a hard time modelling the economy or the dynamics of political systems if they tried to represent the behaviour of every single market participant or every single citizen.

Given the ubiquity of higher-level descriptions in science, some philosophers ask whether the world itself might be "stratified into levels", where different levels are organized hierarchically, perhaps with a fundamental level at the bottom.[4] The levels in question, then, are not just *levels of description* or *explanation*, but *levels of reality* or *ontological levels*. On one view, different descriptive or explanatory levels correspond to different ontological levels: they are "epistemic markers" of something "ontic".

How should we think about levels? Are notions such as *levels of description*, *levels of explanation*, or *ontological levels* mere metaphors, as is sometimes suggested, or can we explicate them precisely? The aim of this paper is to present a general framework for studying levels, whether interpreted epistemically or ontically. I introduce an abstract definition of a *system of levels* and discuss a number of applications, some of which can be interpreted as capturing descriptive or explanatory levels while others can be interpreted as capturing ontological levels. One of these applications captures the idea

[1]See, among many others, Fodor (1974), Owens (1989), and Beckermann, Flohr, and Kim (1992).

[2]For a classic discussion, see Putnam (1967). On levels in cognitive science, see also Bechtel (1994).

[3]See, e.g., Kincaid (1986), Sawyer (2002, 2003), List and Pettit (2011), and List and Spiekermann (2013).

[4]For a defence of the stratified picture, see Schaffer (2003). The quote (de-italicized) is from p. 498.

that a level of description may be a marker of an ontological level. The applications build on some recent discussions of levels in the literature; the underlying abstract definition is inspired by category theory.[5]

I will illustrate the usefulness of the proposed framework by bringing it to bear on some familiar philosophical questions: are levels linearly ordered, and is there a fundamental level?[6] And what does the answer to this question imply for physicalism, the thesis that everything supervenes on (i.e., is determined by) the physical? Are there emergent higher-level properties that are not accompanied by matching lower-level properties? Are higher-level descriptions always reducible to lower-level ones? Can we represent the relationship between normative and non-normative domains as one involving levels? And might a levelled framework shed some light on the relationship between third-personal and first-personal levels, especially on the (often claimed) failure of the first-personal to supervene on the third-personal?[7] My aim is not to give conclusive answers to these questions. It would be preposterous to try to do so within the scope of a single paper. My aim is rather to illustrate how the proposed framework allows us to frame some of the issues in a helpful way.

2 A system of levels: an abstract definition

I begin by giving an abstract definition of a system of levels. In the next section, I discuss some instances of this definition. In some cases, the idea of levels has a more epistemological or explanatory flavour, in others a more ontological one.

A *system of levels* is a pair $\langle \mathcal{L}, \mathcal{S} \rangle$, defined as follows:

- $\mathcal{L}$ is a class of objects called *levels* (which will be given more structure later), and
- $\mathcal{S}$ is a class of mappings between levels, called *supervenience mappings*, where each such mapping σ has a *source level* L and a *target level* L' and is denoted $\sigma : L \to L'$,

such that the following conditions hold:

[5] The closest precursors to the present work are Butterfield (2012), List (2014), and List and Pivato (2015a,b). Himmelreich (2015, Appendix B) also explicates the idea of levels, building on the formalism in List and Pivato (2015a). Category theory goes back to Eilenberg and MacLane (1945). For a philosophical survey, see Marquis (2015). For a recent philosophy-of-science application (specifically, an account of theories as categories), see Halvorson (forthcoming). Category theory has also been suggested as a framework for thinking about levels of description in cognitive and brain science (Gómez Ramirez 2014). However, my proposal here is quite different from those earlier works in the literature.

[6] This is the question discussed in Schaffer (2003).

[7] For arguments against the supervenience of first-personal consciousness on third-personal physical properties, see, e.g., Chalmers (1996, 2004). For an earlier classic contribution, see Nagel (1974).

(S1) $\mathcal{S}$ is closed under *composition* of mappings, i.e., if $\mathcal{S}$ contains $\sigma : L \to L'$ and $\sigma' : L' \to L''$, then it also contains the composite mapping $\sigma \bullet \sigma' : L \to L''$ defined by first applying σ and then applying σ' (where composition is *associative*[8]);

(S2) for each level L, there is an *identity mapping* $\mathbf{1}_L : L \to L$ in $\mathcal{S}$, such that, for every mapping $\sigma : L \to L'$, we have $\mathbf{1}_L \bullet \sigma = \sigma = \sigma \bullet \mathbf{1}_{L'}$;

(S3) for any pair of levels L and L', there is at most one mapping from L to L' in $\mathcal{S}$.

Interpretationally, when the mapping $\sigma : L \to L'$ is contained in $\mathcal{S}$, this means that level L' *supervenes* (or *depends*) on level L. We then call L' the *supervenient* (or *higher*) *level* and L the *subvenient* (or *lower*) *level*, according to σ. Alternatively, we might call σ a *determination mapping*. In philosophy, supervenience is understood as a relation of determination or necessation. One set of facts is said to "supervene" on a second set if the second set of facts determines the first, i.e., a change in the first set of facts is impossible without any change in the second. There can be different notions of supervenience, corresponding to different modes of determination or necessitation; supervenience can be metaphysical or nomological, for example.[9] The formal framework is compatible with different interpretations.

The three conditions on a system of levels capture some familiar properties of the notion of supervenience. Condition (S1) entails that supervenience is *transitive*: if L'' supervenes on L', and L' supervenes on L, then L'' also supervenes on L. Condition (S2) entails that every level supervenes on itself; trivially, supervenience is *reflexive* (though nothing of substance hangs on this). Condition (S3) entails that, whenever L' supervenes on L, the relation in which L and L' stand is *unique*; this is in line with the idea of supervenience as a relation of determination or necessitation. The three conditions jointly entail a fourth condition:

(S4) if $\mathcal{S}$ contains a mapping $\sigma : L \to L'$ and a mapping $\sigma' : L' \to L$, then $\sigma \bullet \sigma' = \mathbf{1}_L$.[10]

Informally, if two levels supervene on one another (which might perhaps never happen if

[8]Formally, $\sigma \bullet (\sigma' \bullet \sigma'') = (\sigma \bullet \sigma') \bullet \sigma''$ whenever $\sigma : L \to L'$, $\sigma' : L' \to L''$, and $\sigma'' : L'' \to L'''$.

[9]One set of facts supervenes metaphysically on a second if a change in the first set of facts is metaphysically impossible without a change in the second. One set of facts supervenes nomologically on a second if this is nomologically impossible, i.e., impossible relative to the appropriate laws of nature.

[10]To see this, suppose that $\mathcal{S}$ contains a mapping $\sigma : L \to L'$ and a mapping $\sigma' : L' \to L$. By (S1), $\mathcal{S}$ will then also contain the composite mapping $\sigma \bullet \sigma' : L \to L$. By (S2), $\mathcal{S}$ contains an identity mapping $\mathbf{1}_L : L \to L$ for level L. By (S3), $\mathcal{S}$ contains at most one mapping from L to L. Since both $\sigma \bullet \sigma'$ and $\mathbf{1}_L$ are mappings from L to L, they must coincide; otherwise $\mathcal{S}$ would contain more than one mapping from L to L. This establishes (S4).

these levels are distinct), then the composite of the relations in which they stand must be the identity relation.

In algebraic terms, the pair $\langle \mathcal{L}, \mathcal{S} \rangle$, subject to conditions (S1) and (S2), is a structure called a "category". Generally, a *category* is a pair consisting of a class of objects and a class of mappings betweens objects, often called "arrows" or "morphisms", where conditions (S1) and (S2) hold. In the present context, the "objects" are levels, and the "arrows" or "morphisms" are supervenience mappings. Categories that also satisfy condition (S3), as in the case of $\langle \mathcal{L}, \mathcal{S} \rangle$, are called "posetal categories".

The category-theoretic way of representing systems of levels allows us to identify structural relationships between different such systems.[11] First of all, one system of levels, $\langle \mathcal{L}, \mathcal{S} \rangle$, is a *subsystem* of another, $\langle \mathcal{L}', \mathcal{S}' \rangle$, if

- $\mathcal{L} \subseteq \mathcal{L}'$ and $\mathcal{S} \subseteq \mathcal{S}'$, and
- composition and identity in $\langle \mathcal{L}, \mathcal{S} \rangle$ are defined as in $\langle \mathcal{L}', \mathcal{S}' \rangle$.[12]

Second, and more generally, there can be structure-preserving mappings between different systems of levels. These are called *functors*. A *functor*, F, from one system of levels, $\langle \mathcal{L}, \mathcal{S} \rangle$, to another, $\langle \mathcal{L}', \mathcal{S}' \rangle$, is a mapping which

- assigns to each level L in $\mathcal{L}$ a corresponding level $L' = F(L)$ in $\mathcal{L}'$, and
- assigns to each supervenience mapping $\sigma : L \to L'$ in $\mathcal{S}$ a corresponding supervenience mapping $\sigma' = F(\sigma)$ in $\mathcal{S}'$, where $\sigma' : F(L) \to F(L')$,

such that F preserves composition and identity.[13] The existence of a functor from one system of levels to another means that we can map the first system into the second in a way that preserves supervenience relationships. If there are functors in both directions (e.g., from $\langle \mathcal{L}, \mathcal{S} \rangle$ to $\langle \mathcal{L}', \mathcal{S}' \rangle$ and from $\langle \mathcal{L}', \mathcal{S}' \rangle$ to $\langle \mathcal{L}, \mathcal{S} \rangle$), where these functors are inverses of one another, this indicates that the two systems of levels are structurally equivalent. The attraction of the present definition of a system of levels is its generality, as I will now illustrate.

[11] Note that, under Gómez Ramirez's (2014) very different proposal, each level – as opposed to a system of levels – is represented by a category (e.g., the category of neurons for the neuronal level, with synaptic paths playing the role of morphisms), and there are no supervenience mappings as morphisms.

[12] Note that $\langle \mathcal{L}, \mathcal{S} \rangle$ and $\langle \mathcal{L}', \mathcal{S}' \rangle$, qua systems of levels, must each satisfy (S1) to (S3).

[13] Formally, for any two supervenience mappings σ and σ' in $\mathcal{S}$, where the target level of σ coincides with the source level of σ', we have $F(\sigma \bullet \sigma') = F(\sigma) \bullet F(\sigma')$; and for any identity mapping $\mathbf{1}_L$ in $\mathcal{S}$, we have $F(\mathbf{1}_L) = \mathbf{1}_{F(L)}$, where $\mathbf{1}_{F(L)}$ is the identity mapping in $\mathcal{S}'$ for level $F(L)$.

3 Four instances of systems of levels

I will discuss four applications of the general definition I have given. Some of these refer explicitly to levels of description, others refer explicitly to ontological levels, and in some cases they admit both interpretations.

3.1 Levels of grain

I begin with a very simple example of a system of levels, which is generated by different ways of partitioning an underlying non-empty set Ω of possible worlds (or other items).[14] Consider an equivalence relation $\sim$ on Ω (a reflexive, symmetrical, and transitive relation). Any such relation $\sim$ partitions Ω into some non-empty, pairwise disjoint, and jointly exhaustive equivalence classes (each of which consists of worlds or items that are equivalent with respect to $\sim$). Let $\Omega_\sim$ denote the resulting set of equivalence classes. We call $\Omega_\sim$ a *partition* of Ω. We obtain the finest partition if $\sim$ is the identity relation; here, each element of Ω forms a singleton equivalence class by itself. We obtain the coarsest partition if $\sim$ is the total relation, under which all elements of Ω fall into the same equivalence class. Non-trivial partitions lie in between these two extremes.

For any two partitions $\Omega_\sim$ and $\Omega_\approx$, we say that $\Omega_\sim$ is *at least as fine-grained* as $\Omega_\approx$ if each equivalence class in $\Omega_\approx$ is a union of equivalence classes in $\Omega_\sim$. The relation "at least as fine-grained as" partially orders partitions. Whenever $\Omega_\sim$ is at least as fine-grained as $\Omega_\approx$, we define a function $\sigma : \Omega_\sim \to \Omega_\approx$ that assigns to each equivalence class in $\Omega_\sim$ the equivalence class in $\Omega_\approx$ in which it is included.

It is easy to see that we get a system of levels if we define the pair $\langle \mathcal{L}, \mathcal{S} \rangle$ as follows:

- $\mathcal{L}$ is some non-empty set of partitions of Ω, perhaps the set of all logically possible partitions;
- $\mathcal{S}$ consists of every function $\sigma : \Omega_\sim \to \Omega_\approx$ under the definition just given, where $\Omega_\sim$ and $\Omega_\approx$ are elements of $\mathcal{L}$ such that $\Omega_\sim$ is at least as fine-grained as $\Omega_\approx$.

A *level*, here, is simply a particular way of partitioning the underlying space of possibilities (worlds or items) into equivalence classes, such that we do not distinguish between members of the same equivalence class. This definition captures what is meant by the *levels of grain* at which we represent the world. The most natural interpretation of this is an epistemic one. Different levels correspond to different ways of perceiving the world.

[14] The idea of identifying levels with partitions is discussed in Himmelreich (2015, Appendix B). Himmelreich develops a version of this idea adapting the framework of worlds-as-histories from List and Pivato (2015a).

In decision theory, for example, an agent's *level of awareness* is often modelled in this way.[15] Someone's *awareness* is defined in terms the distinctions he or she is able to draw. The agent is said to be *aware* of some feature of the world (or a feature of some item) if and only if he or she is able to distinguish worlds (or items) with that feature from ones without it. The more features an agent is aware of, the more distinctions between worlds (or items) he or she is able to draw. Greater awareness thus corresponds to the adoption of a more fine-grained partition of the space of possibilities, lesser awareness to the adoption of a more coarse-grained partition. Accordingly, different levels of awareness can be represented as a system of levels in the present sense.

3.2 Ontological levels

As already noted, it is a familiar idea – though still controversial – that the world itself is stratified into levels.[16] According to this levelled picture, there is not just a single set of possible worlds ("possible worlds *simpliciter*"), but there are different such sets, which encode facts at different levels. Worlds at the physical level encode the totality of physical facts. Worlds at the chemical and biological levels encode the totality of chemical and biological facts. And worlds at the psychological and social levels encode the totality of psychological and social facts.

It is usually assumed that higher levels (which are "more macro") supervene on lower ones (which are "more micro"). For example, the chemical level supervenes on the physical one, insofar as the totality of physical facts determines the totality of chemical facts. Furthermore, higher-level facts are usually assumed to be *multiply realizable* by lower-level facts: different configurations of lower-level facts can correspond to (necessitate, bring about, realize) the same higher-level facts. For instance, many different states of the individual water molecules in a flask can instantiate the same aggregate macro-state of the water. Similarly, a number of subtly different configurations of physical properties can instantiate the same chemical or biological properties.

We can formalize this ontological picture as a system of levels $\langle \mathcal{L}, \mathcal{S} \rangle$, where:

- $\mathcal{L}$ is some non-empty class of sets of "level-specific" worlds (with each set of level-specific worlds non-empty);

[15] See, e.g., Modica and Rustichini (1999). In a recent working paper, Dietrich (2016) has proposed a model of decision-making under uncertainty in which an agent's subjective conceptualization of outcomes and states takes the form of appropriate partitions of some underlying space of possibilities.

[16] See, e.g., Schaffer (2003).

- $\mathcal{S}$ is some class of surjective ("onto") functions of the form $\sigma : \Omega \to \Omega'$,[17] where Ω and Ω' are elements of $\mathcal{L}$, such that $\mathcal{S}$ satisfies (S1), (S2), and (S3).

Each element Ω of $\mathcal{L}$ can be interpreted as an *ontological level*: it is the set of possible worlds at that level. For example, $\mathcal{L}$ might contain a set Ω corresponding to the physical level, a set Ω' corresponding to the chemical level, a set Ω'' corresponding to the biological level, and so on. A physical-level world settles all physical facts; a chemical-level world settles all chemical facts; and so on.

To say that the chemical level supervenes on the physical, or that the biological supervenes on the chemical, is to say that there exists a surjective function $\sigma : \Omega \to \Omega'$, which maps each lower-level world to the higher-level world that it realizes. Surjectivity means that there are no possible worlds at the higher level that lack a lower-level realizer. For example, for a world to be chemically possible – i.e., contained in Ω' – it must have a physical realizer: there must be some $\omega \in \Omega$ such that $\sigma(\omega) = \omega'$. An instance of *multiple realizability* occurs when the function $\sigma : \Omega \to \Omega'$ is "many-to-one": several distinct elements of Ω can realize the same element of Ω'.[18]

We can use the present framework to express not only the idea that higher-level *worlds* in Ω' supervene on lower-level *worlds* in Ω, but also the idea that specific higher-level *facts* supervene on specific lower-level *facts*. Let $E' \subseteq \Omega'$ represent some higher-level fact, namely the fact that the higher-level world falls inside the set E'. We write $\sigma^{-1}(E')$ for the *inverse image* of E' under the supervenience mapping σ, defined as the set of all lower-level worlds that are mapped (by σ) to some element of E', formally $\sigma^{-1}(E') = \{\omega \in \Omega : \sigma(\omega) \in E'\}$. We can then interpret $E = \sigma^{-1}(E')$ as the *supervenience base* of E'. It consists of all the possible lower-level realizers of E'. Whether the higher-level fact E' obtains (i.e., the higher-level world falls inside E') depends on whether the underlying lower-level fact E obtains (i.e., the lower-level world falls inside E). For instance, whether someone is in pain (a psychological fact) supervenes on whether this person's brain is in a pain-generating state, such as "C-fibres firing" (a neuro-physiological fact).

The present understanding of ontological levels differs subtly from one that is common in the philosophical literature. Ontological levels are often understood as *levels of entities and their properties*, where lower-level entities are the building blocks of higher-level entities.[19] The microphysical level, for example, is the level of elementary particles and their properties, while the macrophysical level is the level of larger aggregates. The

[17] A function $\sigma : \Omega \to \Omega'$ is *surjective* ("onto") if, for every $\omega' \in \Omega'$, there exists some $\omega \in \Omega$ such that $\sigma(\omega) = \omega'$.

[18] "Many-to-one" is the negation of injectivity. A function $\sigma : \Omega \to \Omega'$ is *injective* ("one-to-one") if, for any $\omega, \omega' \in \Omega$, $\sigma(\omega) = \sigma(\omega')$ implies $\omega = \omega'$.

[19] See, e.g., Schaffer (2003) and Kim (1993); I quote Kim in Section 4.1 below.

relationship between levels is then a part-whole relationship.[20] Of course, level-specific worlds as I have defined them here can be understood as specifications of level-specific entities and their properties. Nonetheless, I think it is best to define levels primarily in terms of level-specific worlds, and to take level-specific entities only to be derivative. So, I prefer to begin with a specification of level-specific *sets of worlds*, and to treat these only secondarily as specifications of the properties of certain level-specific entities. This picture, which gives primacy to worlds rather than entities, is in line with Wittgenstein's famous dictum:

> "The world is everything that is the case. The world is the totality of facts, not of things."[21]

In a similar vein, we may say: *a level-specific world is everything that is the case at that level; it is the totality of level-specific facts, not of level-specific things.* Of course, if we have a theory of the world at a particular level, where this theory has certain ontological commitments, then we may interpret the entities and properties to which the theory is committed as the level-specific entities and properties.

Finally, note that the levels of grain that I defined in the previous subsection can be formally viewed as a special case of the present definition of ontological levels, although the suggested interpretation of levels of grain was epistemological. If, in the earlier definition, we interpret each coarsened partition of the underlying set Ω as a set of higher-level worlds, then our earlier *levels as partitions* can be re-interpreted as *ontological levels*, as defined in this subsection. Thus, for each system of levels of grain, there exists a structurally equivalent system of ontological levels. (Recall that structural equivalence means that there are functors, in both directions, between the two systems, where these functors are inverses of each other.)

Importantly, the present definition of a system of ontological levels is more general than the earlier definition of levels of grain. Under the present definition, higher-level worlds need not be *identified* with equivalence classes of lower-level worlds; they merely *pick out* such equivalence classes.[22] The definition permits the inclusion in $\mathcal{L}$ of two distinct levels Ω' and Ω'' which each supervene on some lower level Ω and pick out the same equivalence classes of worlds in Ω. In fact, two distinct levels Ω' and Ω'' in $\mathcal{L}$ could

[20] Himmelreich (2015, Appendix B) also distinguishes between a mereological understanding of levels and a world/state-based understanding akin to the one defended here and argues for the latter.

[21] See Wittgenstein (1922, 1 and 1.1).

[22] For any supervenience mapping $\sigma : \Omega \to \Omega'$ in $\mathcal{S}$, each world $\omega' \in \Omega'$ picks out the equivalence class of those worlds $\omega \in \Omega$ such that $\sigma(\omega) = \omega'$.

supervene on one another and could thus be viewed as distinct but isomorphic.[23] By contrast, if levels are *partitions* of some underlying set of worlds, no two distinct levels of grain could ever supervene on one another.

Although a system of ontological levels is formally more general than a system of levels of grain, there exists a functor from any system of ontological levels to some system of levels of grain. That system will then mirror some (though not necessarily all) of the structure of our system of ontological levels. We can arrive at this functor in two steps. First, we must identify a lowest level, i.e., a level on which all levels supervene. If the system $\langle \mathcal{L}, \mathcal{S} \rangle$ already has a lowest level, then this step is straightforward. But if it has no lowest level – a possibility to which we return in Section 4.1 – we must construct a hypothetical level on which all levels supervene, a so-called "inverse limit". For a posetal category such as $\langle \mathcal{L}, \mathcal{S} \rangle$, the construction of an inverse limit is possible. (Formally, this involves extending $\langle \mathcal{L}, \mathcal{S} \rangle$ to a larger category which contains this inverse limit. Of course, we need not interpret it as anything more than a mathematical construct.) In the second step, we can associate each level in $\mathcal{L}$ with the partition of "lowest-level" worlds that the given ontological level picks out. In this way, we can map the given system of ontological levels to some system of levels of grain.[24]

3.3 Levels of description

Regardless of whether we consider a levelled ontology independently plausible, it is undeniable that we use different levels of description to think and speak about the world. In fundamental physics, we describe the world in different terms than in the special sciences, such as chemistry, biology, psychology, or the social sciences. And within each of these sciences, there are debates about which level of description is appropriate for the phenomena of interest: the level of individual molecules versus that of larger aggregates in physics and chemistry, the level of the cell versus that of the organism or ecosystem in biology, the level of the brain versus that of the mind in psychology, and the level of individuals versus that of larger social entities in the social sciences. The notion of a level of explanation is closely related to that of a level of description. An explanation at a particular level – say, a macroeconomic explanation – is an explanation that uses descriptions at that level.

[23]Formally, this means that $\mathcal{S}$ contains a supervenience mapping $\sigma : \Omega \to \Omega'$ and also a supervenience mapping $\sigma' : \Omega' \to \Omega$. It follows from our definitions that each of these mappings must then be *bijective* (i.e., injective and surjective). Surjectivity follows from the definition of $\mathcal{S}$. If injectivity were violated, we would not have $\sigma \bullet \sigma' = \mathbf{1}_L$, thereby contradicting condition (S4) in the previous section.

[24]I am indebted to Marcus Pivato for suggesting the inverse-limit construction.

To define a system of levels of description, I begin by introducing the notion of a *language* that we may use to talk about the world.[25] I define a *language*, $\mathbf{L}$, as a set of formal expressions – called *sentences* – which is endowed with two things:

- a *negation operator*, denoted $\neg$, such that, for each sentence $\phi \in \mathbf{L}$, there exists a corresponding negated sentence, $\neg\phi \in \mathbf{L}$;
- a notion of *consistency*, which deems some sets of sentences consistent and the remaining sets of sentences inconsistent.[26]

An example of a language is standard propositional logic. Here $\mathbf{L}$ is the set of all well-formed sentences that can be constructed out of some atomic sentences and the standard logical connectives ("and", "or", "not", "if-then", and so on), and we call a set of sentences *consistent* if all its members can be simultaneously true. Other examples of languages are more expressive logics, such as predicate, modal, conditional, and deontic logics. A Boolean algebra can also qualify as a language in the present sense, where this algebra is a set $\mathcal{A}$ of subsets of some underlying set Ω of possible worlds (with Ω non-empty), such that $\mathcal{A}$ is closed under intersection, union, and complementation. A standard example is the set of *all* subsets of Ω. Here the role of "sentences" is played by elements of $\mathcal{A}$. Any set of such elements is *consistent* if these elements have a non-empty intersection. (Recall that each element of $\mathcal{A}$ is a subset of Ω.)

Crucially, any language $\mathbf{L}$, as I have defined it, induces a corresponding "ontology", understood as a minimally rich set of worlds $\Omega_{\mathbf{L}}$ such that each world in $\Omega_{\mathbf{L}}$ "settles" everything that can be expressed in $\mathbf{L}$. To *settle* a sentence is to assign a determinate truth-value to it: either "true" or "false". If one takes the sentences in $\mathbf{L}$ to have truth-conditions, then one is, in effect, committed to positing such an ontology. The set $\Omega_{\mathbf{L}}$ can be interpreted as the set of all possible ways the world could be such that

(i) everything that is expressible in $\mathbf{L}$ is settled, and

(ii) nothing else is settled that is not entailed by what is expressible in $\mathbf{L}$.

One cannot take the language $\mathbf{L}$ at face value (i.e., be a realist about the contents expressible in it) without assuming that there is a fact of the matter as to which element

[25] I borrow the present abstract definition of a language from Dietrich (2007), who introduced it in a different context, namely that of judgment-aggregation theory.

[26] The notion of consistency must satisfy three minimal conditions (Dietrich 2007): (i) any sentence-negation pair $\{\phi, \neg\phi\}$ is inconsistent; (ii) any superset of any inconsistent set is inconsistent; (iii) the empty set is consistent and every consistent set has a consistent superset containing a member of each sentence-negation pair in $\mathbf{L}$.

of $\Omega_{\mathbf{L}}$ is the actual one. In that sense, the language $\mathbf{L}$ is a "marker" of the associated ontology $\Omega_{\mathbf{L}}$.

For modelling purposes, the easiest way to define the set $\Omega_{\mathbf{L}}$ is to take it to be the set of all maximally consistent subsets of $\mathbf{L}$. A *maximally consistent* subset of $\mathbf{L}$ is a consistent set of sentences to which no further sentences can be added without undermining consistency. Alternatively, if taking worlds to *be* maximally consistent subsets of $\mathbf{L}$ is too artificial, we only need to assume that the worlds in $\Omega_{\mathbf{L}}$ *correspond* to the maximally consistent subsets of $\mathbf{L}$.[27] We say that a sentence $\phi \in \mathbf{L}$ is *true* at a world $\omega \in \Omega_{\mathbf{L}}$ if the maximally consistent subset of $\mathbf{L}$ to which ω corresponds contains ϕ; the sentence is *false* otherwise. For each sentence $\phi \in \mathbf{L}$, we write $[\phi]$ to denote the set of worlds in $\Omega_{\mathbf{L}}$ at which ϕ is true; we call this the *extension* of ϕ.[28]

We call any pair consisting of a language $\mathbf{L}$ and the induced set of worlds $\Omega_{\mathbf{L}}$ a *level of description*. We can now define a system of levels of description $\langle \mathcal{L}, \mathcal{S} \rangle$ as follows:

- $\mathcal{L}$ is some non-empty class of levels of description, each of which is a pair $\langle \mathbf{L}, \Omega_{\mathbf{L}} \rangle$;
- $\mathcal{S}$ is some class of surjective functions of the form $\sigma : \Omega_{\mathbf{L}} \to \Omega_{\mathbf{L}'}$, where $\langle \mathbf{L}, \Omega_{\mathbf{L}} \rangle$ and $\langle \mathbf{L}', \Omega_{\mathbf{L}'} \rangle$ are levels of description in $\mathcal{L}$, such that $\mathcal{S}$ satisfies (S1), (S2), and (S3).

For example, $\mathcal{L}$ may contain levels corresponding to fundamental physics, chemistry, biology, psychology, and the social sciences. Each such level is a pair of an appropriate level-specific language and the induced set of level-specific worlds. The supervenience mappings capture the idea that chemical-level worlds supervene on physical-level worlds, biological-level worlds supervene on chemical-level worlds, and so on. In this way, a system of levels of description can capture the different levels corresponding to the different special sciences; the supervenience mappings between them capture the relationships between levels. I return to those relationships in Section 4.3, where I discuss whether supervenience entails reducibility, in a sense to be made precise.[29]

[27] Formally, there exists a bijection from $\Omega_{\mathbf{L}}$ to the set of maximally consistent subsets of $\mathbf{L}$.

[28] The set of all such extensions, $\{[\phi] : \phi \in \mathbf{L}\}$, is structurally equivalent to our language $\mathbf{L}$.

[29] It is worth mentioning one important special case of a system of levels of description. Here, each pair $\langle \mathbf{L}, \Omega_{\mathbf{L}} \rangle$ in $\mathcal{L}$ is of the following form: $\Omega_{\mathbf{L}}$ is some partition $\Omega_{\sim}$ of an underlying non-empty set Ω of possible worlds as in Section 3.1 (where $\sim$ is the equivalence relation generating that partition), and $\mathbf{L}$ is some canonical algebra $\mathcal{A}_{\Omega_{\sim}}$ over $\Omega_{\sim}$ (in the simplest case, the set of all subsets of $\Omega_{\sim}$). We can then define the supervenience mappings in $\mathcal{S}$ as in Section 3.1, i.e., two levels $\langle \mathcal{A}_{\Omega_{\sim}}, \Omega_{\sim} \rangle$ and $\langle \mathcal{A}_{\Omega_{\approx}}, \Omega_{\approx} \rangle$ are related by some mapping $\sigma : \Omega_{\sim} \to \Omega_{\approx}$ if and only if $\Omega_{\sim}$ is at least as fine-grained as $\Omega_{\approx}$. A notable feature of this case is that whenever there is such a mapping between two levels, the higher-level algebra $\mathcal{A}_{\Omega_{\approx}}$ is (isomorphic to) a subalgebra of the lower-level algebra $\mathcal{A}_{\Omega_{\sim}}$. In particular, each element of $\mathcal{A}_{\Omega_{\approx}}$ has an inverse image (with respect to σ) in $\mathcal{A}_{\Omega_{\sim}}$. As will become clear in Section 4.3, this feature is not shared by levels of description in general.

Note that by focusing just on the sets of worlds induced by each level-specific language, we can map a system of levels of description to a corresponding system of ontological levels, as defined earlier. Technically, there is a functor from any system of levels of description to the induced system of ontological levels. However, systems of levels of description are structurally richer than systems of ontological levels, by encoding descriptions as well as ontologies. Different systems of levels of description could induce structurally equivalent systems of ontological levels.

3.4 Levels of dynamics

Ever since the development of statistical mechanics, there has been considerable interest in the dynamics of physical and other systems at different levels (or "scales"). A coin-tossing system can be studied at a microphysical level, where the focus is on the precise details of the coin's trajectory as it is being tossed. Alternatively, the system can be studied at a statistical-mechanical level, where the coin is viewed as a simple Bernoulli-distributed stochastic process with only two possible outcomes, "heads" or "tails". Similarly, the weather, climate, or the economy can each be studied at a micro-level, where the focus is on detailed processes at a fine-grained resolution, or at a macro-level, where the system is specified in terms of certain aggregate variables. To mark this contrast, we often speak of a system's "higher-level dynamics" and its "lower-level dynamics". Practically any interesting dynamic system can be studied at multiple levels, and as we will see later, the dynamic properties of such a system – for instance, whether it is deterministic or not – may depend on the level in question.[30] I will now briefly explain how such "levels of dynamics" fit into the present framework.

I begin with a simple definition of a *temporally evolving system*.[31] Let T be some linearly ordered set of points of *time*, where "$<$" stands for the "before" relation. Let X be some set of *states* in which the system can be at any time; we call X the system's *state space*. A *history* of the system is a function from time into the space space, $h : T \to X$, which assigns to each time $t \in T$ the system's state at that time, denoted $h(t)$. We write Ω to denote the set of all histories that are permitted by the laws of the system. We can think of these histories as the *nomologically possible* ones. The set Ω plays the role of the set of possible worlds. Collections of histories are called *events*.

We further require the notion of a *conditional probability structure*. This is a family of conditional probability functions, $\{Pr_E\}_{E \subseteq \Omega}$, which contains one conditional probability

[30] See, e.g., Butterfield (2012), List and Pivato (2015a), and Werndl (2009).

[31] The definition of a temporally evolving system and its analysis at different levels are based on List and Pivato (2015a,b). For a more basic analysis without a specification of probabilities, see List (2014).

function Pr_E for each event $E \subseteq \Omega$. Each Pr_E assigns to any event $F \subseteq \Omega$ the conditional probability of that event, given E, denoted $Pr_E(F)$. For example, to determine the conditional probability of some event F in history h at time t, we need to conditionalize on the event that we have reached time t in history h; so, the probability in question is $Pr_E(F)$, where E is the set of all histories $h' \in \Omega$ that coincide with h up to time t. We call the pair $\langle \Omega, \{Pr_E\}_{E \subseteq \Omega} \rangle$ a *temporally evolving system.*

To see how we can study temporally evolving systems at multiple levels, let such a system be given, and interpret its state space X as a set of lower-level states (e.g., microphysical states of some coin-tossing system). Now assume that each state in X determines (realizes, instantiates) some higher-level state in some other set X': a *higher-level state space* (e.g., a set of aggregate states such as "heads" or "tails"). For each higher-level state (an element X'), there is an equivalence class of lower-level states (a subset of X) that may realize that higher-level state. For instance, each micro-state of the billions of individual water molecules in a flask determines a corresponding macro-state of the water. Let $\sigma : X \to X'$ be the function that assigns to each lower-level state the resulting higher-level state.

We call σ a *supervenience mapping* from the given lower-level system $\langle \Omega, \{Pr_E\}_{E \subseteq \Omega} \rangle$ to a resulting higher-level system $\left\langle \Omega', \{Pr'_{E'}\}_{E' \subseteq \Omega'} \right\rangle$ if it has the following properties:

(i) every higher-level state in X' has at least one possible lower-level realizing state in X according to σ, where X and X' are the two systems' state spaces; formally, the function $\sigma : X \to X'$ is surjective;

(ii) the set Ω determines the set Ω' via σ; formally, σ induces a surjective mapping from Ω to Ω': for each history $h \in \Omega$, $\sigma(h) = h'$, where, for each $t \in T$, $h'(t) = \sigma(h(t))$;

(iii) the probability structure $\{Pr_E\}_{E \subseteq \Omega}$ determines the probability structure $\{Pr'_{E'}\}_{E' \subseteq \Omega'}$ via σ; formally, for any pair of events $E', F' \subseteq \Omega'$, $Pr'_{E'}(F') = Pr_E(F)$, where E and F are the inverse images of E' and F' under σ, respectively.[32]

We are now in a position to define a system of levels, $\langle \mathcal{L}, \mathcal{S} \rangle$:

- $\mathcal{L}$ is some non-empty class of "level-specific" temporally evolving systems, where, for each set Ω, $\mathcal{L}$ contains at most one system $\langle \Omega, \{Pr_E\}_{E \subseteq \Omega} \rangle$ whose set of histories is Ω (i.e., each Ω is endowed with a unique conditional probability structure);

- $\mathcal{S}$ is some class of supervenience mappings with the properties just defined, such that $\mathcal{S}$ satisfies (S1), (S2), and (S3).

[32] In line with the previous definition, the *inverse image* of any $E' \subseteq \Omega'$ under σ is $\{h \in \Omega : \sigma(h) \in E'\}$.

So, when two level-specific temporally evolving systems in $\mathcal{L}$ are related via a supervenience mapping σ in $\mathcal{S}$, the dynamics of the higher-level system is determined by the dynamics of the lower-level system: higher-level states (in X') supervene on lower-level states (in X), and higher-level histories as well as higher-level probabilities are determined by lower-level ones.[33] We are then able to consider how the dynamics at different levels relate to one another; I return to this issue in Section 4.2.

Again, the present example of a system of levels can be related to one of our earlier examples. By focusing just on the sets of level-specific histories, where each history is interpreted as a possible world, we can map any system of level-specific dynamic systems to a corresponding system of ontological levels, as defined in Section 3.2. We have thereby constructed a functor from a system of levels of the present kind to one of the earlier kind. Of course, level-specific probabilistic information is lost under this functor. In that sense, a system of level-specific dynamic systems is structurally richer than a system of ontological levels under the earlier definition.

4 Some illustrative philosophical questions

As we have seen, the present framework can capture a variety of instances of systems of levels, where levels may be interpreted either as levels of description or as ontological levels, and in some cases in both ways. I will now show how the framework can be brought to bear on some familiar philosophical questions. As already noted, my aim is not to give conclusive answers to these questions, but merely to suggest that the framework can be used to approach the relevant debates in a helpful way.

4.1 Is there a linear hierarchy of levels, with a fundamental level at the bottom?

A positive answer to this question is widely assumed, but seldom carefully defended, as Jonathan Schaffer notes. He quotes Jaegwon Kim at length:

> "The Cartesian model of a *bifurcated* world has been replaced by that of a *layered* world, a hierarchically stratified structure of 'levels' or 'orders' of entities and their characteristic properties. It is generally thought that there is a bottom level, one consisting of whatever microphysics is going to tell us are the most basic physical particles out of which all matter is composed (electrons, neutrons, quarks, or whatever)."[34]

[33] As noted in List and Pivato (2015b), the higher-level system is a *factor system* of the lower-level one.
[34] See Kim (1993, p. 337), quoted in Schaffer (2003).

Similarly, David Owens writes:

> "[T]he levels metaphor naturally suggests itself as a way of visualizing the structure of science. According to this picture, there is a hierarchy made up of different levels of explanation. Physics is at the base of this hierarchy and the rest of the structure depends upon it."[35]

Are we justified in assuming that there is a linearly ordered hierarchy of this kind, with a fundamental level at the bottom? The first thing to note is that the question of whether levels are linearly ordered is distinct from the question of whether there is a fundamental level at the bottom. A system of levels $\langle \mathcal{L}, \mathcal{S} \rangle$ is *linear* if the levels in $\mathcal{L}$ are totally ordered by the supervenience mappings, i.e., supervenience is transitive, antisymmetric, and complete.[36] The system $\langle \mathcal{L}, \mathcal{S} \rangle$ *has a fundamental level* if there is some level in $\mathcal{L}$ on which every level supervenes. The abstract definition of a system of levels that I have given does not automatically secure either of these properties.

First consider linearity. A system of levels is, in general, only partially ordered.[37] The relation of supervenience is transitive (as well as reflexive), but not generally antisymmetric or complete. While systems of levels of grain are antisymmetric by definition (in that no two distinct levels of grain can mutually supervene on each other), systems of the other kinds (e.g., systems of ontological levels or levels of description) need not be antisymmetric: two distinct levels could supervene on one another. Further, completeness need not be satisfied either. It is not generally the case that, for any two levels L and L', either L supervenes on L' or vice versa (or both). Only special cases of systems of levels are totally ordered. Even if we define levels in the simplest way, as levels of grain, the resulting system of levels is only partially ordered by the relation "at least as fine-grained as". Although the terminology of "levels" is conventional, the alternative terminology of "scales" captures the lack of a linear hierarchy: there can be many different scales, which need not form a single hierarchy.

A partially ordered system of levels, in turn, could include more than one linearly ordered chain, where different such chains meet at most in a few places. For example, the system of levels could look like a tree, or even an upside-down tree, where some levels supervene on – or subvene – others that do not themselves stand in any supervenience relation relative to each other. For example, a scenario in which, for each level L, there

[35] See Owens (1989, p. 59).

[36] I have already defined *transitivity*. *Antisymmetry* means that if L supervenes on L' and L' also supervenes on L, then $L = L'$. *Completeness* means that, for any two levels L and L', L supervenes on L' or L' supervenes on L.

[37] For a similar observation, see also List and Pivato (2015a, fn. 41).

are two distinct levels L' and L'' such that L supervenes on each of them is entirely coherent. If levels are given by level-specific sets of worlds, then $L = \Omega$ might supervene on each of $L' = \Omega \times \Phi$ and $L'' = \Omega \times \Psi$, where Φ and Ψ are disjoint. Here, each world at level L' is an ordered pair consisting of an element of Ω and an element of Φ, and each world at level L'' is an ordered pair consisting of an element of Ω and an element of Ψ. The supervenience mappings from L' to L and from L'' to L then map each L'-level world or each L''-level world to their first component. Figure 1 illustrates such a scenario.

Figure 1: A non-linear system of levels

$$\begin{array}{cccc} & \cdots & & \\ & \uparrow & & \\ & L & & \\ & \nearrow \quad \nwarrow & & \\ L' & & & L'' \\ \nearrow \quad \nwarrow & & & \nearrow \quad \nwarrow \\ \cdots \quad \cdots & & & \cdots \quad \cdots \end{array}$$

Next consider the question of a fundamental level. In general, there need not exist a level on which every level supervenes. If there were a fundamental level, say with a set Ω of "bottom-level" worlds, then any world in Ω would determine not only all facts at the fundamental level, but also all higher-level facts, via the appropriate supervenience mappings. We might then interpret an element of Ω as a "world *simpliciter*". However, a partially ordered system of levels could have multiple distinct lowest levels: one for each linearly ordered chain. Furthermore, a system of levels, whether totally ordered or not, could have infinitely descending chains, in which each level supervenes on an even lower level, as in the example of Figure 1. Technically, a system of levels need not be *well-founded*. This observation is consistent with Jonathan Schaffer's conclusion that there is no evidence in favour of the assumption that there is a fundamental level. Schaffer himself defends a "metaphysic of infinite descent".[38] The present framework confirms the coherence of such a scenario.[39]

[38]See Schaffer (2003, p. 499).

[39]For an earlier discussion of this point, see List and Pivato (2015a). As noted in Section 3.2 above, for a system of ontological levels, where each level is a set of level-specific worlds, it is possible to construct an inverse limit. So, any system of ontological levels without a fundamental level can be mathematically viewed as a subsystem of a larger system in which there is a fundamental level. Of course, this larger

What does this imply for physicalism, the thesis that everything supervenes on the physical? The exact meaning of "physicalism" depends, among other things, on how we define "the physical".[40] We could either define "the physical" in terms of our *current* best account of what the fundamental physical facts are. Or, alternatively, we could define "the physical" in terms of the best future account of those facts, whatever it turns out to be. Irrespective of the definition we adopt, however, the thesis that everything supervenes on the physical can be true only if there is *some* level on which all other levels supervene, i.e., a fundamental level. If there is no such level, physicalism is a non-starter. At best, physicalism could be true as a claim about a certain subclass of levels – those that do in fact supervene on the physical level. Perhaps the meteorological and chemical levels supervene on the physical one, for example. But as an all-encompassing thesis, physicalism would be structurally false if a particular kind of bottomless levelled ontology were vindicated. Alternatively, we could define a weaker notion of physicalism, according to which some subclass of levels, say $\mathcal{L}_{\text{phys}} \subseteq \mathcal{L}$, is designated as "physical", where every level in $\mathcal{L}$ supervenes on some level in $\mathcal{L}_{\text{phys}}$. This, however, is a watered-down notion of physicalism.

It should be clear, then, that the question of whether or not there is a fundamental level has far-reaching philosophical consequences. Since – conceptually speaking – a system of levels need not have a fundamental level, physicalism seems to stand on weaker ground than usually acknowledged.

4.2 Are there emergent properties?

For present purposes, a *level-specific property* is a property that may be instantiated by a world or object at a particular level. A *higher-level property*, in particular, is a property that may be instantiated by a world or object at a higher level. The notion of *emergence* refers to the idea that some properties may be instantiated at some higher level without being simultaneously instantiated at any lower level. Emergence is consistent with supervenience. We may say that a higher-level property is *emergent* if it supervenes on lower-level properties but is not generally accompanied by some corresponding ("type-equivalent") lower-level property. By contrast, we may say that a higher-level property is *matched at the lower level* if it is always accompanied by a corresponding lower-level property. Of course, these informal definitions can be made more precise.

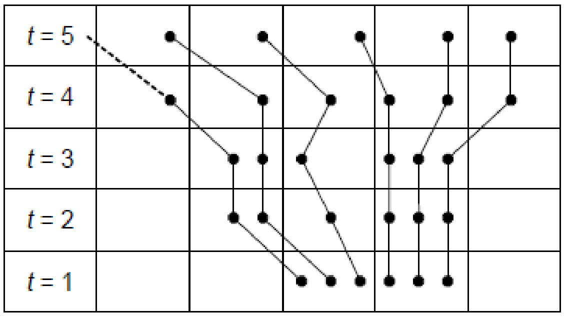
As there is a sizeable literature on the topic of emergence, I will review only a single

system could be a mere mathematical construct.

[40] On the notion of physicalism, see, e.g., Stoljar (2010).

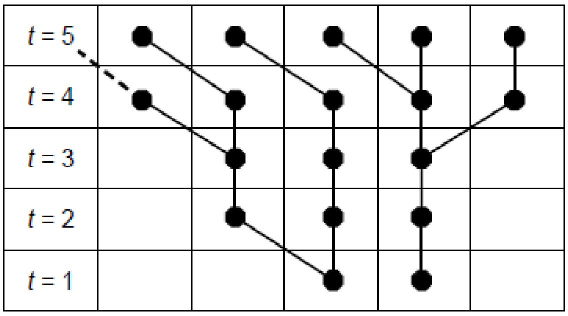
example, namely that of emergent indeterminism.[41] Consider a temporally evolving system (as defined in Section 3.4) whose histories across five time periods, $t = 1, 2, ..., 5$, are as shown in Figure 2.[42] Each dot represents one state in the state space X, and each path through the figure from bottom to top represents one possible history. The set of

Figure 2: Lower-level histories



(reproduced from List 2014)

Figure 3: Higher-level histories



(reproduced from List 2014)

all such paths is Ω. Clearly, all histories in Ω are *deterministic*, in the sense that any initial segment of any history admits only a single possible continuation in Ω.[43]

Now consider the temporally evolving system at a higher, more macroscopic level. We interpret the original states in X and the histories in Ω as lower-level histories,

[41] For earlier analyses of this phenomenon, see, e.g., Butterfield (2012), List (2014), List and Pivato (2015a), and Werndl (2009).

[42] The present example comes from List (2014).

[43] Formally, for any history $h : T \to X$, the *initial segment* of h up to time t, denoted h_t, is the restriction of the function h to all points in time up to t. History h' is a *continuation* of an initial segment h_t if $h'_t = h_t$. History h is *deterministic* if every initial segment of h has only one continution in Ω, namely h itself. History h is *indeterministic* if some initial segment of h has more than one continuation in Ω.

and introduce a higher-level state space X' that results from X via a many-to-one supervenience mapping. Specifically, suppose that any lower-level states that lie in same rectangular box in Figure 2 realize the same higher-level state. So, the supervenience mapping treats all lower-level states within the same box in the grid as belonging to the same equivalence class. Figure 3 shows the image of Figure 2 under this supervenience mapping. The thick dots represent higher-level states, and the possible paths through the figure from bottom to top represent possible histories, this time at the higher level. Note that the set Ω' of higher-level histories is the image of Ω under the given supervenience mapping. Clearly, higher-level histories are *indeterministic* in this example: the initial segment of any higher-level history admits more than one possible continuation in Ω'.[44]

The example shows that indeterminism may be an emergent property. A temporally evolving system may display indeterminism at a higher level, consistently with determinism at a lower level. One may not even be able to ask meaningfully whether a system is deterministic or indeterministic *simpliciter*. The answer to this question depends entirely on the level at which we are considering the system.[45] Determinism and indeterminism are examples of level-specific properties.

Needless to say, there are many other real-world instances of level-specific properties. An example is the property of being in a particular intentional state, such as believing or desiring something. This can only be instantiated by a person at a psychological level. A person's physical organism does not have that property. Indeed, it would be a category mistake to employ that property at a lower level: it does not belong to the vocabulary of fundamental physics or even to that of neuro-physiology. Similarly, a particular unemployment rate is a higher-level property; this can only be instantiated by a society or an economy, but not by its physical supervenience base.

4.3 Are higher-level descriptions reducible to lower-level ones?

It is often assumed that because higher-level phenomena supervene on lower-level ones, we should also be able to *explain* all higher-level phenomena in terms of lower-level ones. The idea, in short, is that supervenience implies explanatory reducibility. Because chemical phenomena supervene on physical ones, for example, we should ultimately be able to explain chemical phenomena in terms of physical ones. Similarly, because social phenomena supervene on the interactions of a large number of individuals, we should ultimately be able to explain social phenomena in terms of individual-level processes. If

[44] Formally, determinism and indeterminism are defined as they are at the lower level, except that the quantification is now over higher-level histories (in Ω') rather than lower-level histories (in Ω).

[45] For earlier arguments for this claim, see List and Pivato (2015a) and relatedly Werndl (2009).

those reductionistic claims were true, then higher-level descriptions would be dispensable for many purposes. They might be nothing more than shorthand notations for certain things that we could equally express at a lower level.

However, as several philosophers have argued in relation to domains ranging from the philosophy of mind to the philosophy of social science, supervenience does not imply explanatory reducibility.[46] If we were to dispense with higher-level descriptions, we would run the risk of overlooking some higher-level patterns in the world.[47] Systematic regularities in the world are not confined to the physical level or some other lower level. Rather, they can occur at higher levels too. For example, the causes that make a counterfactual difference to some effect may sometimes be certain higher-level properties, rather than their lower-level realizers.[48] The difference-making cause of a decrease in inflation may be the increase in the interest rate by the central bank – a higher-level property – rather than its individual-level or physical-level realizers. While inflation may systematically co-vary with the interest rate, it need not equally systematically co-vary with any particular lower-level realizing properties on which the interest rate supervenes.

The present framework lends further support to the claim that higher-level descriptions may be irreducible to lower-level ones, even if there is a fundamental level on which all other levels supervene. (If there is no fundamental level, then *all* levels are higher levels relative to some other levels, and so the attempt to "reduce away" higher-level descriptions could not get off the ground.) Consider a system of levels of description, as introduced in Section 3.3, and assume that there is a fundamental level at the bottom, given by the pair $\langle \mathbf{L}, \Omega_{\mathbf{L}} \rangle$, where $\mathbf{L}$ is the level-specific language – say that of fundamental physics – and $\Omega_{\mathbf{L}}$ is the level-specific set of worlds. Let us make two assumptions. First, $\Omega_{\mathbf{L}}$ is an infinite set. This is a reasonable assumption; plausibly, infinitely many worlds are compatible with the fundamental laws of nature, provided that we allow variations in initial conditions. Second, the language $\mathbf{L}$ is countable; i.e., it admits infinitely many expressions but no more than there are natural numbers. This is also a reasonable assumption; all familiar languages are countable, from propositional logic to English.

Now let us turn to some higher level of description, given by the pair $\langle \mathbf{L}', \Omega_{\mathbf{L}'} \rangle$, where $\mathbf{L}'$ is some higher-level language and $\Omega_{\mathbf{L}'}$ is the corresponding set of higher-level worlds. Let σ be the supervenience mapping from the fundamental level to the present higher level; formally, this is a surjective function of the form $\sigma : \Omega_{\mathbf{L}} \to \Omega_{\mathbf{L}'}$. We want to know whether all higher-level descriptions are "reducible" to corresponding lower-level

[46]See, e.g., Putnam (1967), Fodor (1974), Owens (1989), Sawyer (2002, 2003), List and Pettit (2011), List and Spiekermann (2013), and Dietrich and List (2016, Section 8).

[47]See, e.g., Dennett (1991).

[48]See, e.g., List and Menzies (2009, 2010). On higher-level causation, see also Glynn (2013).

descriptions, so that we would then be able to translate higher-level explanations into lower-level ones. Call a higher-level sentence $\phi' \in \mathbf{L}'$ *reducible* to a lower-level sentence from $\mathbf{L}$ if it is extensionally equivalent (modulo supervenience) to some sentence in $\mathbf{L}$: there is some sentence $\phi \in \mathbf{L}$ whose extension (in $\Omega_{\mathbf{L}}$) is the inverse image, under σ, of the extension (in $\Omega_{\mathbf{L}'}$) of ϕ'; formally, $[\phi] = \sigma^{-1}([\phi'])$. It is important to note that, if the language $\mathbf{L}$ is closed under conjunction and disjunction, as in the case of standard propositional logic, then reducibility to a single sentence is equivalent to reducibility to any finite logical combination of sentences (because any such logical combination can be expressed as a single composite sentence). If every higher-level sentence were reducible to a corresponding lower-level sentence (or a finite combination of lower-level sentences), then we might indeed consider higher-level descriptions dispensable, at least in principle. Higher-level explanations would be translatable into lower-level ones.

However, there is an important combinatorial reason why reducibility is the exception rather than the rule, as I will now explain.[49] The supervenience of the higher level on the lower level entails that:

(i) the inverse image, under σ, of the extension of any higher-level sentence is some subset of $\Omega_{\mathbf{L}}$.

However, it does not follow that:

(ii) this subset can be described by some sentence (or by a finite combination of sentences) from the lower-level language $\mathbf{L}$.

For reducibility, both (i) and (ii) are needed; (i) alone is not enough. Why might (ii) fail? The set $\Omega_{\mathbf{L}}$, being infinite, has uncountably many subsets, of which $\mathbf{L}$ allows us to describe only countably many. Since $\mathbf{L}$ is countable, the set of subsets of $\Omega_{\mathbf{L}}$ that can be finitely described in $\mathbf{L}$ is also countable. Therefore, almost all subsets of $\Omega_{\mathbf{L}}$ – namely all except a countable number – do *not* admit a finite description in $\mathbf{L}$. For almost every subset of $\Omega_{\mathbf{L}}$, then, there exists no sentence in $\mathbf{L}$ (or even a finite combination of sentences) whose extension is that subset. Of course, it it is logically possible that two levels of description are so well aligned that all higher-level sentences can be reduced to equivalent lower-level sentences, in that the inverse images of their extensions are the extensions of some lower-level sentences. But, from a combinatorial perspective, this is a highly special case. It would be surprising – a "cosmic coincidence" – if the levels corresponding to different special sciences turned out to be like this.

The bottom line is that, even when there is a supervenience relation between a lower level and a higher one, higher-level descriptions are not generally reducible to

[49] The present argument draws on List and Pivato (2015a, Section 8).

lower-level desciptions. And even in those special cases in which a higher-level sentence *can* be reduced to some lower-level sentence (or a finite combination of sentences), this translation may well be so cumbersome and uneconomical as to be of little practical use: for instance, the lower-level sentences might be unmanageably long. So, higher-level descriptions may be irreducible in practice, even if they are not irreducible in principle.

4.4 Can we represent the is-ought relationship in a levelled framework?

There is much discussion about the relationship between normative and non-normative domains of discourse.[50] How does the normative domain relate to the non-normative one (the "descriptive" domain)? I will here take the normative domain to be represented by language involving "obligation" and "permission" operators ("ought" and "may"), while I will take the non-normative domain to be represented by language that is free from such operators. For the sake of argument, I will assume that the sentences we express in normative language can be true or false. That is: there are truth-conditions for the normative domain, just as there are truth-conditions for the descriptive domain. Some people, notably non-cognitivists, reject this assumption. Even among those who accept the existence of truth-conditions for the normative domain, there is little agreement on what those truth-conditions are or on how they relate to the truth-conditions for the descriptive domain. Naturalists, for instance, think that normative truths supervene on non-normative ones, while non-naturalists disagree. I suggest that we can helpfully think about the relationship between the normative and non-normative domains by representing them as two different levels of description, with their associated ontologies.

Let $\mathbf{L}$ denote some descriptive, non-normative language, and let $\Omega_{\mathbf{L}}$ be the associated set of worlds. We can think of $\mathbf{L}$ as our non-normative "base language", and we can think of the worlds in $\Omega_{\mathbf{L}}$ as fully specifying all descriptive facts. We now augment the language by introducing normative terms. For present purposes, these will be the deontic operators "it is obligatory that" and "it is permissible that". I assume that these operators have the standard properties assumed in deontic logic. (My arguments could equally be developed if the normative language were specified differently.) Let $\mathbf{L}^+$ be the "normatively augmented" language. While the original language $\mathbf{L}$ can express only descriptive discourse, the augmented language $\mathbf{L}^+$ can express both descriptive and normative discourse. I suggest that we can think of $\mathbf{L}$ and $\mathbf{L}^+$ as giving rise to two different levels of description.

Our central question is this: what set $\Omega_{\mathbf{L}^+}$ of worlds is associated with $\mathbf{L}^+$, and how does it relate to $\Omega_{\mathbf{L}}$, the set of worlds associated with $\mathbf{L}$? In other words, how is

[50] For a recent discussion of the is-ought gap, see Brown (2014).

the normative level, $\langle \mathbf{L}^+, \Omega_{\mathbf{L}^+} \rangle$, related to the non-normative one, $\langle \mathbf{L}, \Omega_{\mathbf{L}} \rangle$? In particular, is there a supervenience mapping $\sigma : \Omega_{\mathbf{L}} \to \Omega_{\mathbf{L}^+}$ and/or a supervenience mapping $\sigma^+ : \Omega_{\mathbf{L}^+} \to \Omega_{\mathbf{L}}$? I will argue that there exists a supervenience mapping from $\Omega_{\mathbf{L}^+}$ to $\Omega_{\mathbf{L}}$, but that the converse holds only in special cases.

The first thing to note is that since $\mathbf{L}^+$ is an augmented version of $\mathbf{L}$, every sentence from $\mathbf{L}$ is also contained in $\mathbf{L}^+$. And so, since any world in $\Omega_{\mathbf{L}^+}$ settles all sentences from $\mathbf{L}^+$, it must also settle all sentences from $\mathbf{L}$. This shows that, to each world ω^+ in $\Omega_{\mathbf{L}^+}$, there corresponds a world ω in $\Omega_{\mathbf{L}}$ such that ω^+ and ω assign the same truth-values to all sentences from $\mathbf{L}$. Let σ^+ be the function that maps each $\omega^+ \in \Omega_{\mathbf{L}^+}$ to the corresponding $\omega \in \Omega_{\mathbf{L}}$. If we assume that any consistent subset of $\mathbf{L}$ remains consistent when viewed as a subset of $\mathbf{L}^+$, it follows further that, for every world $\omega \in \Omega_{\mathbf{L}}$, there exists at least one world $\omega^+ \in \Omega_{\mathbf{L}^+}$ with $\sigma^+(\omega^+) = \omega$. Therefore σ^+ is surjective. This establishes the existence of a supervenience mapping $\sigma^+ : \Omega_{\mathbf{L}^+} \to \Omega_{\mathbf{L}}$.

What about the converse? The sentences in $\mathbf{L}^+$, unlike those in $\mathbf{L}$, may involve the operators "it is obligatory that" and "it is permissible that", abbreviated "O" and "P". In line with standard deontic logic, we assume that $\mathrm{O}\phi$ is true at some world if and only if ϕ is true at *all* worlds that are permissible relative to that world. The permissible worlds, in turn, are specified by a *selection function*, f, which assigns to each world a set of *permissible worlds* relative to the given world. Similarly, $\mathrm{P}\phi$ is true at some world if and only if ϕ is true at *some* world that is permissible relative to it. But over which worlds should we quantify in this definition? Should we quantify over the worlds in $\Omega_{\mathbf{L}^+}$ or over the worlds in $\Omega_{\mathbf{L}}$?

Clearly, if we take the selection function f as given, we do not need to quantify over worlds from $\Omega_{\mathbf{L}^+}$. Given f, all the information needed to define the truth-values of $\mathrm{O}\phi$ and $\mathrm{P}\phi$ is already encoded in the "non-normative" worlds from $\Omega_{\mathbf{L}}$. So, any world $\omega \in \Omega_{\mathbf{L}}$, together with the selection function f, settles all the sentences from $\mathbf{L}^+$, including those that go beyond $\mathbf{L}$. In other words, if the selection function f is held fixed in the background, we can identify $\Omega_{\mathbf{L}^+}$ with $\Omega_{\mathbf{L}}$, and, by implication, there is a supervenience mapping $\sigma : \Omega_{\mathbf{L}} \to \Omega_{\mathbf{L}^+}$. We can then think of $\mathbf{L}^+$ and $\mathbf{L}$ as two different languages whose associated ontologies are essentially the same, despite the fact that $\mathbf{L}^+$ is expressively richer than $\mathbf{L}$. On this picture, normative and non-normative forms of discourse refer to the same world, just described differently. This, I think, is the picture that proponents of naturalism about normativity have in mind.

By contrast, if we do not take the selection function f as given, the supervenience of the normative level on the non-normative one breaks down. Without the function f, the worlds in $\Omega_{\mathbf{L}}$ are insufficient to settle everything that can be expressed in $\mathbf{L}^+$. The truth-values of sentences such as Oϕ and Pϕ are left open. So, worlds in $\Omega_{\mathbf{L}}$ encode strictly less information than worlds in $\Omega_{\mathbf{L}^+}$. Note that, for something to qualify as a world in $\Omega_{\mathbf{L}^+}$, it must settle everything that can be expressed in $\mathbf{L}^+$.

Let F be some set of admissible selection functions. In order to settle everything that can be expressed in $\mathbf{L}^+$, we require not only a world $\omega \in \Omega_{\mathbf{L}}$, but also a selection function $f \in F$. We may thus define $\Omega_{\mathbf{L}^+}$ as the set of all possible pairs of the form $\langle \omega, f \rangle$, formally $\Omega_{\mathbf{L}^+} = \Omega_{\mathbf{L}} \times F$ (or a suitable subset thereof). Any world in $\Omega_{\mathbf{L}^+}$, formally a pair $\langle \omega, f \rangle$, will then suffice to settle all sentences in $\mathbf{L}^+$. As before, there is a supervenience mapping $\sigma^+ : \Omega_{\mathbf{L}^+} \to \Omega_{\mathbf{L}}$, but the converse is no longer generally true. The mapping from $\Omega_{\mathbf{L}^+}$ to $\Omega_{\mathbf{L}}$ can be many-to-one. However, there is one important special case in which there exists a supervenience mapping $\sigma : \Omega_{\mathbf{L}} \to \Omega_{\mathbf{L}^+}$. This is the case in which F is singleton: there exists only a single admissible selection function. In this case, $\Omega_{\mathbf{L}^+}$ and $\Omega_{\mathbf{L}}$ are isomorphic (so that they can essentially be identified with one another), which is again the picture that naturalists have in mind.

Recall that there are different notions of supervenience. Earlier I gave the examples of metaphysical and nomological supervenience. We can now observe the following: *relative to a given selection function*, the normative level does indeed supervene on the non-normative one. This is a form of *relativized supervenience*, i.e., supervenience relative to some further parameter, here the selection function f. Without such a relativization, the normative level does not supervene on the non-normative one. Perhaps the hidden assumption underlying naturalism is just a particular choice of selection function.

4.5 What is the relationship between third-personal and first-personal levels?

As conscious subjects, we are not merely biological organisms that function in certain ways and can be studied from some external perspective, but we ourselves experience the world from a first-person perspective. There is something it is like to be a conscious subject, as Thomas Nagel puts it.[51] By contrast, many entities and systems in the world, including some fairly complex ones, have no conscious experiences. The weather system, an eco-system, the global economy, and a smartphone are each systems of considerable complexity, and yet – for all we know – there is nothing it is like to be such a system. We can study these systems from the outside – from a third-personal perspective – but they have no "inner life": there is no first-personal perspective attached to them.

How does consciousness fit into our scientific worldview? David Chalmers describes

[51] See Nagel (1974).

the explanatory challenge as follows:

> "[T]he distinctive task of a science of consciousness is to systematically integrate two key classes of data into a scientific framework: third-person data about behavior and brain processes, and first-person data about subjective experience. When a conscious system is observed from the third-person point of view, a range of specific behavioral and neural phenomena present themselves. When a conscious system is observed from the first-person point of view, a range of specific subjective phenomena present themselves... [B]oth sorts of phenomena have the status of data for a science of consciousness."[52]

Crucially, Chalmers argues, first-personal data cannot be explained solely in terms of third-personal data; first-personal data are, in an important sense, "irreducible". This echoes Nagel's earlier point:

> "[First-personal experience] is not captured by any of the familiar ... reductive analyses of the mental, for all of them are logically compatible with its absence. It is not analyzable in terms of any explanatory system of functional states, or intentional states, since these could be ascribed to robots or automata that behaved liked people though they experienced nothing."[53]

Some philosophers, such as Daniel Dennett, deny that first-personal experience is irreducible to third-personal phenomena or that it needs to be explained at all.[54] But, arguably, such a view fails to do justice to the phenomenon of conscious experience. I suggest that we can helpfully think about the relationship between first-personal and third-personal phenomena by viewing them as phenomena on different levels. To explain this point in a simple way, I will use the variant of my framework in which levels are identified with level-specific sets of worlds. (The point could also be developed in a more sophisticated way, but this is beyond the scope of this paper.)

Let me begin with the third-personal level. We can represent this level as a set of possible worlds that are specified as richly as necessary to capture all third-personal facts. Call this set Ω_{3rd}. Depending on whether we are interested in metaphysical or nomological possibility, the elements of Ω_{3rd} could be either all metaphysically possible third-personal worlds or alternatively all nomologically possible ones, i.e., those worlds that are compatible with the laws of nature. The present framework permits both interpretations. (Furthermore, strictly speaking, there is not just one third-personal

[52] See Chalmers (2004).
[53] See Nagel (1974, p. 436).
[54] See, e.g., Dennett (2005).

level, but there are many, corresponding to the different special sciences, as already discussed. However, if Ω_{3rd} corresponds to the fundamental physical level, we might think of the other third-personal levels as supervening on that level. So, we can take Ω_{3rd} to correspond to the bottom level within some partially ordered (sub)system of third-personal levels – assuming, for the sake of argument, that such a level exists.)

Any world ω in Ω_{3rd} encodes all third-personal facts. For example, if there are biological organisms in the world, ω encodes all facts about their brains, bodies, and behaviour, as well as all facts about their environment. At the same time, ω does not encode any first-personal facts. How exactly a subject's first-personal perspective is related to ω, or indeed whether there is such a perspective at all, is left open. David Lewis makes an analogous point as follows:

> "Consider the case of the two gods. They inhabit a certain possible world, and they know exactly which world it is. Therefore they know every proposition that is true at their world. Insofar as knowledge is a propositional attitude [with third-personal content], they are omniscient. Still I can imagine them to suffer ignorance: neither one knows which of the two he is."[55]

Put differently, a third-personal world does not specify the *place* of any first-personal subject inside that world – the "I". Even an omniscient being, as in Lewis's thought experiment, would not be able to infer his or her own first-personal perspective from the totality of third-personal facts, if those were the only facts to which he or she had access. The first-personal subject is absent from a world such as ω. It is then reasonable to conclude that the first-personal facts are not determined by the third-personal facts.

In order to place a subject inside the world, we need to specify something above and beyond the third-personal world ω, namely a "locus of subjectivity".[56] Call this π. This encodes a subject's first-personal perspective on ω; the letter "pi" stands for "perspective". A first-personal world, then, is a pair $\langle \omega, \pi \rangle$, consisting of a third-personal world and a perspective on it. Formally, this is analogous to what philosophers call a *centred world*: a world paired with some location or "centre".[57] Yet this notion is typically understood more narrowly than what is required here: centres are often defined as spatio-temporal locations within a world, akin to the dot indicating your current location on your smartphone map. By contrast, I interpret a "locus of subjectivity" more broadly, namely as encoding a subject's entire first-personal perspective on ω, however

[55]See Lewis (1979, p. 520). Lewis's main point is subtly different from mine (his topic is indexical belief, not first-personal experience), but the lesson of his quote carries over to my discussion.

[56]This echoes some ideas from phenomenology. For a survey, see, e.g., Gallagher and Zahavi (2015).

[57]Centred worlds go back to Quine (1969) and Lewis (1979). For a helpful discussion, see Liao (2012).

richly this may need to be specified. It may include, for instance, a full specification of the subject's phenomenal experience.[58]

The pair $\langle \omega, \pi \rangle$ encodes the totality of facts about the world ω, *with* π *placed inside it as the locus of subjectivity.* By definition, this subsumes both third-personal and first-personal facts. To be clear: to say that the pair $\langle \omega, \pi \rangle$ is "my" first-personal world is not to say that I *know* all the facts specified by $\langle \omega, \pi \rangle$. Rather, the pair $\langle \omega, \pi \rangle$ captures everything that is *true* about the world that I inhabit, with myself as the locus of subjectivity. I may be – and typically will be – ignorant of many of those truths.

We can identify the first-personal level with the set of all possible first-personal worlds. Call this set Ω_{1st}. Formally, it is some subset of $\Omega_{\text{3rd}} \times \Pi$, the set of all possible pairs $\langle \omega, \pi \rangle$ with $\omega \in \Omega_{\text{3rd}}$ and $\pi \in \Pi$, where Π is some universal set of "loci of subjectivity". Again, we can understand Ω_{1st} either as the set of all metaphysically possible first-personal worlds or as the set of all nomologically possible such worlds. Depending on the interpretation, Ω_{1st} may or may not coincide with the entire set $\Omega_{\text{3rd}} \times \Pi$. Plausibly, only certain specific loci of subjectivity are compatible with each third-personal world, where "compatibility", in turn, might be understood metaphysically or nomologically. For example, it might be that a locus of subjectivity must be associated with an entity with the appropriate consciousness-supporting make-up, such as a living, non-comatose, and non-sleeping organism with a brain. Or perhaps, many more loci of subjectivity are compatible with each third-personal world. On a panpsychist view, for instance, a very large number of first-personal perspectives may each be compatible with a given third-personal world. On such a view, even some very simple systems may have their own "perspective". In short, we might have different philosophical and/or scientific views about how exactly the third-personal facts constrain the available first-personal perspectives.[59] We can define a *psychophysical law* as a specification of

[58] Chalmers (1996, p. 144) notes that a centred world, in the standard narrow sense, would be insufficient to capture a subject's entire first-personal perspective, including his or her phenomenal experience. He writes: "indexicals [whose propositional content may be represented by a set of centred worlds] accompany facts about conscious experience in their failure to supervene logically on physical facts, but they are all settled by the addition of a thin 'indexical fact' about the location of the agent in question. But even when we give [the agent] perfect knowledge about her indexical relation to everything in the physical world, her knowledge of [e.g.] red experiences will not be improved in the slightest. In lacking phenomenal knowledge, she lacks far more than someone lacking indexical knowledge." These observations underline the need to adopt a broad interpretation of a locus of subjectivity.

[59] For instance, integrated information theory, a recent popular theory of consciousness, would suggest that first-personal perspectives must always be associated with loci of maximal informational integration in a system (for an overview, see Tononi and Koch 2015). According to this theory, then, a first-personal perspective π is compatible with a third-personal world ω if and only if π is appropriately associated

which pairs of the form $\langle \omega, \pi \rangle$ are included in Ω_{1st} and which are not.

Once we represent the third-personal and first-personal levels in this way – namely in terms of the level-specific sets of worlds Ω_{3rd} and Ω_{1st} – we can analyze their relationship formally. The first thing to note is this. Since first-personal worlds, as I have defined them, are richer than third-personal worlds, there is a supervenience mapping from Ω_{1st} to Ω_{3rd}. Thus, the third-personal level supervenes on the first-personal one. At first, this may sound counterintuitive: an apparent claim about the primacy of the subjective. However, I am not saying that third-personal worlds supervene on "pure" first-personal perspectives; i.e., I am not saying that there is supervenience mapping from the set Π of perspectives to the set Ω_{3rd} of third-personal worlds. Rather, the mapping is from the set Ω_{1st} to the set Ω_{3rd}, and although I have called the elements of Ω_{1st} "first-personal worlds", we can think of them as "first-personally enriched" worlds. Under this interpretation, it is not surprising that the third-personal worlds, being less rich, can supervene on them.

Does there also exist a converse supervenience mapping, i.e., from Ω_{3rd} to Ω_{1st}? Materialists typically give an affirmative answer, dualists a negative one.[60] To compare these answers in the present framework, consider first the materialist answer. If there is a supervenience mapping from from Ω_{3rd} to Ω_{1st}, then – since there is also a supervenience mapping from Ω_{1st} to Ω_{3rd} – the sets Ω_{3rd} and Ω_{1st} must stand in a one-to-one correspondence: the existence of supervenience mappings in both directions entails that there is exactly one first-personal world corresponding to each third-personal world.[61] This implies, in particular, that only one locus of subjectivity is compatible with each third-personal world. But is this plausible?

Perhaps – and this is very speculative – if we could somehow hold the subject fixed, for example by focusing just on myself, then there might be just one way in which this particular subject – my own "I" – can be placed in the world. For example, it might be nomologically impossible for this particular "I" to be associated with any physical substrate distinct from my actual biological body and brain. So, *relative to myself as a subject*, each third-personal world ω might be consistent with only one perspective π. If so, my own first-personal world $\langle \omega, \pi \rangle$ would supervene on the third-personal world ω, *in this relativized sense of supervenience*. Formally, for each ω, the set Ω_{1st} would contain

with some locus of maximal informational integration in ω. A functionally awake human cortex is an example of such a locus, while an ecosystem or a current-generation smartphone is not.

[60]Here, supervenience is usually understood metaphysically. Chalmers (1996) suggests that consciousness may supervene nomologically on physical properties, relative to some psychophysical law.

[61]As noted before, if there is a supervenience mapping $\sigma : \Omega_{3rd} \to \Omega_{1st}$ and also a supervenience mapping $\sigma' : \Omega_{1st} \to \Omega_{3rd}$, then both of these mappings must be bijective.

a unique pair $\langle \omega, \pi \rangle$ with ω as the first component. We would then have to interpret Ω_{1st} as the set of all possible first-personal worlds *for the given fixed subject.*

However, as soon as we recognize that there are many subjects that can each have their own first-personal perspective on the same third-personal world ω – not just myself, but also you and every other conscious being – it becomes implausible to assume that each third-personal world ω can be paired only with a single locus of subjectivity. Rather, there are many different possible first-personal worlds of the form $\langle \omega, \pi \rangle$, which share the same third-personal world as a component. We can think of these first-personal worlds as different "first-personal realizers" of the same third-personal world. If this is right, then first-personal and third-personal worlds stand in a many-to-one relationship, rather than a one-to-one relationship; and only the third-personal level supervenes on the first-personal, not the other way round, contrary to the materialist picture.

We can observe something else: since different subjects, by definition, "inhabit" different first-personal worlds, there is no such thing as the first-personal world *simpliciter,* i.e., a first-personal world that we all share: you, I, and everyone else. Rather, as conscious subjects, we all live in different first-personal worlds – in effect, "parallel" ones – which are distinct first-personal-level realizers of the same shared third-personal world. Of course, our physical organisms can be thought of as entities within the same shared third-personal world. We can think of the present picture as a "many-worlds model of consciousness".[62] For each subject, there is an *actual first-personal world*, but this is different for different subjects. There is also an *actual third-personal world*: the one that is realized by all those actual first-personal worlds. Just as higher-level worlds are often multiply realizable by lower-level worlds, so third-personal worlds are multiply realizable by first-personal worlds.

The present framework also allows us to address the much-discussed topic of zombies, on which materialists and dualists disagree.[63] A *zombie* is a hypothetical being which

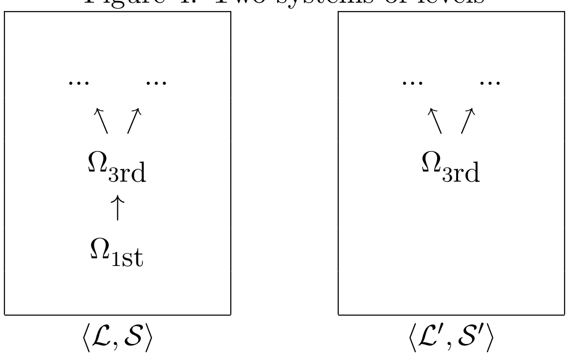
[62] The idea of "many worlds" is sometimes invoked in relation to consciousness. Chalmers (1996, ch. 10) suggests that his dualistic theory may be combined with Everett's "many-worlds" interpretation of quantum mechanics. However, he distinguishes between the "splitting-worlds" variant of the Everett view (a genuine "many-worlds" interpretation, which he rejects as a misinterpretation of Everett) and the "one-big-world" variant (which he prefers). According to the latter, "[t]here is only one world, but it has more in it than we might have thought" (p. 347). Specifically, each conscious mind "perceives a separate discrete world, corresponding to the sort of world that we perceive – call this a *miniworld*, as opposed to the *maxiworld* of the superposition. The real world is a maxiworld, and the miniworlds are merely in the minds of the subjects" (ibid.). There would then still be a single world underlying all the different first-personal (mini)worlds. This differs from the "many-worlds" model I have sketched, but in my framework the first-personal level could also supervene on an even more fundamental level.

[63] For an overview, see, e.g., Chalmers (1996).

is physically and functionally indistinguishable from a human being like you or me, but which nonetheless lacks any first-personal consciousness. A zombie has the exact same third-personal properties as its conscious counterpart: its bodily make-up is the same, as is its brain-functioning; it behaves and speaks in the same way as you or I do. It just lacks first-personal consciousness. While a conscious being has a conscious perspective on the world, the zombie does not. Now, importantly, no-one in the mainstream debate suggests that there are zombies in the real world. Rather, the debate concerns the question of whether the notion of a zombie is coherent: are zombies metaphysically possible or not? Dualists answer in the affirmative, materialists in the negative. Let me use the term "zombie scenario" to refer to a scenario in which the world is physically or third-personally indistinguishable from the actual world but in which there is no first-personal consciousness. Is this scenario coherent?

The framework I have sketched suggests that it is. I have argued that we can model consciousness by recognizing that there is a first-personal level in addition to the third-personal one, where the first-personal level *subvenes* the third-personal one, rather than *supervenes* on it. (Recall that other, "higher" third-personal levels, such as those corresponding to various special sciences, may supervene on the lowest third-personal level.) Call this system of levels $\langle \mathcal{L}, \mathcal{S} \rangle$. Then both Ω_{3rd} and Ω_{1st} are among the levels in $\mathcal{L}$ (as may be other, higher third-personal levels). Further, a mapping of the form $\sigma : \Omega_{1st} \to \Omega_{3rd}$ is among the supervenience mappings in $\mathcal{S}$ (as may be other supervenience mappings with higher levels as their targets). But now consider a different system of levels, to be called $\langle \mathcal{L}', \mathcal{S}' \rangle$, which is just like $\langle \mathcal{L}, \mathcal{S} \rangle$, except that the level Ω_{1st} and all supervenience mappings with source-level Ω_{1st} (such as $\sigma : \Omega_{1st} \to \Omega_{3rd}$) are removed. Clearly, $\langle \mathcal{L}', \mathcal{S}' \rangle$ is a subsystem of $\langle \mathcal{L}, \mathcal{S} \rangle$ and qualifies as a perfectly coherent system of levels. Figure 4 schematically illustrates the systems $\langle \mathcal{L}, \mathcal{S} \rangle$ and $\langle \mathcal{L}', \mathcal{S}' \rangle$. Crucially,

Figure 4: Two systems of levels



with respect to all levels above and including Ω_{3rd}, $\langle \mathcal{L}', \mathcal{S}' \rangle$ is indistinguishable from

$\langle \mathcal{L}, \mathcal{S} \rangle$. The two systems differ only with respect to the lowest level: while $\langle \mathcal{L}, \mathcal{S} \rangle$ has a first-personal level at the bottom, $\langle \mathcal{L}', \mathcal{S}' \rangle$ does not.[64] I believe that $\langle \mathcal{L}', \mathcal{S}' \rangle$ represents the zombie scenario and illustrates its coherence.

Finally, my analysis suggests that it is not really meaningful to speak of "a world in which there are zombies", if by "world" we mean "third-personal world". Whether or not there are zombies depends, not on the features of any particular world in Ω_{3rd}, but rather on whether the third-personal worlds in Ω_{3rd} are underwritten, or realized, by first-personal worlds in Ω_{1st}. By definition, no features of a third-personal world could allow us to distinguish between zombies and non-zombies. Indeed, the third-personal worlds in the system $\langle \mathcal{L}', \mathcal{S}' \rangle$ are indistinguishable from those in the system $\langle \mathcal{L}, \mathcal{S} \rangle$. On the other hand, once we step inside any first-personal world, there is, by definition, a subject in that world: a first-personal world is a world with a subjective perspective. So, no properties of a first-personal world could mark the distinction between zombies and non-zombies either. For this reason, the debate about zombies is best interpreted as a debate about which levels are available, not as a debate about what properties there are in a given level-specific world. I conclude that the question of how first-personal consciousness fits into a scientific worldview is really a question about the status of a particular level: the first-personal one.

5 Concluding remarks

I have given an abstract definition of a system of levels, inspired by ideas from category theory, and I have discussed several applications of this definition, which show that the proposed framework can capture notions such as levels of description, levels of explanation, and ontological levels. I have further suggested that the framework can be helpfully brought to bear on some familiar philosophical questions, ranging from questions about the existence of a fundamental level, the defensibility of physicalism, and the (ir)reducibility of higher-level descriptions to questions about the relationship between normative and non-normative domains and between first-personal and third-personal phenomena. Since references to "levels" are ubiquitous in science and philosophy, I hope that, by going beyond metaphorical invocations of this idea and offering a precise formalization, the present framework will have further useful applications.

[64] There is a theoretical possibility of embedding $\langle \mathcal{L}, \mathcal{S} \rangle$ in an even richer system of levels in which the first-personal level supervenes on some even lower level, but I do not explore this possibility here.

6 References

Bechtel, William (1994) "Levels of Description and Explanation in Cognitive Science", *Minds and Machines* 4(1): 1-25.

Beckermann, Ansgar, Hans Flohr, and Jaegwon Kim (eds.) (1992) *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*, Berlin (de Gruyter).

Brown, Campbell (2014) "Minding the Is-Ought Gap", *Journal of Philosophical Logic* 43(1): 53-69.

Butterfield, Jeremy (2012) "Laws, Causation and Dynamics at Different Levels", *Interface Focus* 2(1): 101-114.

Chalmers, David (1996) *The Conscious Mind*, New York (Oxford University Press).

Chalmers, David (2004) "How Can We Construct a Science of Consciousness?" in Michael S. Gazzaniga (ed.), *The Cognitive Neurosciences III*, third edition, Cambridge, MA (MIT Press), pp. 1111-1120.

Dennett, Daniel (1991) "Real Patterns", *Journal of Philosophy* 88(1): 27-51.

Dennett, Daniel (2005) *Sweet Dreams: Philosophical Obstacles to a Scientific Theory of Consciousness*, Cambridge, MA (MIT Press).

Dietrich, Franz (2007) "A generalised model of judgment aggregation", *Social Choice and Welfare* 28(4): 529-565.

Dietrich, Franz (2016) "Decisions under uncertainty with variable concepts of outcomes and states", working paper, Paris School of Economics.

Dietrich, Franz, and Christian List (2016) "Mentalism versus behaviourism in economics: A philosophy-of-science perspective", *Economics and Philosophy* (in press).

Eilenberg, Samuel, and Saunders MacLane (1945) "General Theory of Natural Equivalences", *Transactions of the American Mathematical Society* 58(2): 231-294.

Fodor, Jerry (1974) "Special Sciences (or: The Disunity of Science as a Working Hypothesis)", *Synthese* 28(2): 97-115.

Gallagher, Shaun and Dan Zahavi (2015) "Phenomenological Approaches to Self-Consciousness", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), available at: http://plato.stanford.edu/archives/spr2015/entries/self-consciousness-phenomenological/

Glynn, Luke (2013) "Is There High-Level Causation?", PhilSci-Archive, available at: http://philsci-archive.pitt.edu/9823/

Gómez Ramirez, Jaime (2014) *A New Foundation for Representation in Cognitive and Brain Science: Category Theory and the Hippocampus*, Dordrecht / Heidelberg (Springer).

Halvorson, Hans (forthcoming) "Scientic Theories", in Paul Humphreys (ed.), *The Oxford Handbook of Philosophy of Science*, Oxford (Oxford University Press).

Himmelreich, Johannes (2015) *Agency as difference-making: Causal Foundations of Moral Responsibility*, PhD thesis, London School of Economics.

Kim, Jaegwon (1993) "The nonreductivist's troubles with mental causation", in *Supervenience and Mind: Selected Philosophical Essays*, Cambridge (Cambridge University Press), pp. 336-357.

Kincaid, Harold (1986) "Reduction, explanation, and individualism", *Philosophy of Science* 53(4): 492-513.

Lewis, David (1979) "Attitudes De Dicto and De Se", *The Philosophical Review* 88(4): 513-543.

Liao, Shen-yi (2012) "What Are Centered Worlds?", *The Philosophical Quarterly* 62(247): 294-316.

List, Christian (2014) "Free will, determinism, and the possibility of doing otherwise", *Noûs* 48(1): 156-178.

List, Christian, and Peter Menzies (2009) "Non-reductive physicalism and the limits of the exclusion principle", *Journal of Philosophy* CVI(9): 475-502.

List, Christian, and Philip Pettit (2011) *Group Agency: The Possibility, Design, and Status of Corporate Agents*, Oxford (Oxford University Press).

List, Christian, and Marcus Pivato (2015a) "Emergent Chance", *The Philosophical Review* 124(1): 119-152.

List, Christian, and Marcus Pivato (2015b) "Dynamic and stochastic systems as a framework for metaphysics and the philosophy of science", working paper.

List, Christian, and Kai Spiekermann (2013) "Methodological Individualism and Holism in Political Science: A Reconciliation", *American Political Science Review* 107(4): 629-643.

Marquis, Jean-Pierre (2015) "Category Theory", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), available at: http://plato.stanford.edu/archives/win2015/entries/category-theory/

Menzies, Peter, and Christian List (2010) "The Causal Autonomy of the Special Sciences", in Cynthia Mcdonald and Graham Mcdonald (eds.), *Emergence in Mind*, Oxford (Oxford University Press), pp. 108-128.

Modica, Salvatore, and Aldo Rustichini (1999) "Unawareness and Partitional Information Structures", *Games and Economic Behavior* 27(2): 265-298.

Nagel, Thomas (1974) "What Is It Like to Be a Bat?" *Philosophical Review* 83(4): 435-450.

Owens, David (1989) "Levels of Explanation", *Mind* 98(389): 59-79.

Putnam, Hilary (1967) "Psychological predicates", in W. H. Capitan and D. D. Merrill (eds.), *Art, Mind, and Religion*, Pittsburgh (Univ. of Pittsburgh Press), pp. 37-48.

Quine, W.V. (1969) "Propositional objects", in *Ontological Relativity and Other Essays*, New York (Columbia University Press), pp. 139-160.

Sawyer, R. Keith (2002) "Nonreductive Individualism. Part I – Supervenience and Wild Disjunction", *Philosophy of the Social Sciences* 32(4): 537-559.

Sawyer, R. Keith (2003) "Nonreductive Individualism. Part II – Social Causation", *Philosophy of the Social Sciences* 33(2): 203-224.

Schaffer, Jonathan (2003) "Is there a fundamental level?" *Noûs* 37(3): 498-517.

Stoljar, Daniel (2010) *Physicalism*, London (Routledge).

Tononi, Giulio, and Christof Koch (2015) "Consciousness: here, there and everywhere?" *Philosophical Transactions of the Royal Society B* 370(1668).

Werndl, Charlotte (2009) "Are deterministic descriptions and indeterministic descriptions observationally equivalent?" *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 40(3): 232-242.

Wittgenstein, Ludwig (1922) *Tractatus Logico-Philosophicus*, London (Kegan Paul).