Alan C. Love and Andreas Hüttemann

Comparing Part-Whole Reductive Explanations in Biology and Physics[1]

Abstract

Many biologists and philosophers have worried that importing models of reasoning from the physical sciences obscures our understanding of reasoning in the life sciences. In this paper we discuss one example that partially validates this concern: part-whole reductive explanations. Biology and physics tend to incorporate different models of temporality in part-whole reductive explanations. This results from differential emphases on compositional and causal facets of reductive explanations, which have not been distinguished reliably in prior philosophical analyses. Keeping these two facets distinct facilitates the identification of two further aspects of reductive explanation: intrinsicality and fundamentality. Our account provides resources for discriminating between different types of reductive explanation and suggests a new approach to comprehending similarities and differences in the explanatory reasoning found in biology and physics.

1. Biology, Physics, and Nagel's Reductionist Shadow

Ernest Nagel's philosophical account of theory reduction in the sciences has cast a long shadow on discussions of the relationship between biology and physics.[2] In addition to debates among philosophers of science, some biologists, such as Ernst Mayr, took Nagel's account as the epitome of what is wrong with philosophical analyses of science; namely, a Procrustean maneuver that stretched and hacked biology to fit a conception of science forged primarily on exemplars from physics.

1 Both authors contributed equally to this paper. We are grateful for the comments and suggestions we have received on this material from many colleagues, including Ingo Brigandt, Tom Doyle, Susan Hawthorne, Marie Kaiser, Peter McLaughlin, Ken Schaffner, Ken Waters, and Marcel Weber. Useful feedback also came from participants at the 2009 workshop 'Explanation, Confirmation, and Prediction in Biology and Medicine,' held in Konstanz, Germany and sponsored by the European Science Foundation (Research Networking Programme). We want to express our appreciation for financial support from Alexander von Humboldt Foundation to pursue this collaboration, which grew out of our mutual participation in the Second German-American Frontiers of Humanities Symposium, Hamburg, Germany, October 2005, sponsored by the Alexander von Humboldt Foundation and the American Philosophical Society.

2 Nagel (1961)

D. Dieks et al. (eds.), *Explanation, Prediction, and Confirmation*, The Philosophy of Science in a European Perspective 2, DOI 10.1007/978-94-007-1180-8_13,

> One of my special concerns has been the neglect of biology in works claiming to be philosophies of science. From the 1920s to the 1960s the logical positivists and physicalists who dominated philosophy of science had little interest in and even less understanding of biology, because it simply did not fit their methodology.[3]

Mayr thought this reflected a general tendency among both scientists and philosophers: "The assumption that it should be possible to "reduce" the theories and concepts of all other sciences, including biology, to those of the physical sciences has clearly dominated not only philosophy but science itself" (1). Unsurprisingly, Mayr saw these as failed maneuvers: "Attempts to 'reduce' biological systems to the level of simply physico-chemical processes have failed because during the reduction the systems lost their specifically biological properties" (1). His refrain is a common one—biological systems have special or complex properties, sometimes labeled 'emergent,' which cannot be handled purely in physical or chemical terms: "Where organisms differ from inanimate matter is in the organization of their systems" (2). But Mayr also thought other differences between physics and biology could be relevant, including physical science preferences for single cause explanations and deterministic models.

For Mayr, this was no mere intellectual skirmish. As Director of the Harvard Museum of Comparative Zoology and a public spokesmen for evolutionary biology in the 1960s, Mayr was involved in major battles over resources with the increasingly dominant molecular approaches to biology (populated by expatriate physicists) that flourished after the structure of DNA was discovered.[4] It was in this context that Mayr marshaled his distinction between proximate and ultimate causation, the latter capturing a distinct place for evolutionary biology that was untouchable by molecular biologists.[5] A reorientation of physics-dominated philosophy of science was more than academic; it meant the life (or death) of evolutionary biology. Thus, it is no surprise that Mayr's framing of the issue is in terms of 'autonomy': is biology an autonomous science? What makes biology unique?[6]

Mayr drew a distinction between three different forms of reduction[7]:

(i) constitutive: any dissection of phenomena, events, and processes into the constituents of which they are composed
(ii) explanatory: claims that all the phenomena and processes at higher hierarchical levels can be explained in terms of the actions and interactions of the components at the lowest hierarchical levels

3 Mayr (1988, p. 1)
4 Beatty (1994)
5 Mayr (1961)
6 cf. Mayr (2004)
7 Mayr (1988, pp. 10-11)

(iii) theory: the theories and laws formulated in biology are only special cases of theories and laws formulated in the physical sciences, and that such biological theories can thus be reduced to physical sciences

Mayr took the first to be uncontroversial, while the latter two were areas of contention. Post-positivist philosophers, in his estimation, had decided against theory reduction (Nagel's heritage being the preeminent example) and he argued against explanatory reduction in terms of the complexity of living systems, their possession of a genetic program, the incommensurability between concepts in biology and physics, the different role played by laws, and the presence of ultimate causation in biology. In an effort to stress the failure of explanatory reduction, Mayr pointed to the paleontologist George Simpson's "reverse reduction" of physics to biology.

> The point is that *all* known material processes and explanatory principles apply to organisms, while only a limited number of them apply to nonliving systems. … Biology, then, is the science that stands at the center of all science, and it is here, in the field where all the principles of all the sciences are embodied, that science can truly become unified.[8]

Although Mayr acknowledged that Simpson might have overstated his case, the conclusion that any explanatory reduction of biology to physics was a failure could not be missed.

Perhaps as a consequence of his real-time, real-world battle over reductionism, Mayr overlooked multiple nuances in Nagel's original discussion. In particular, Nagel was quite sensitive to the possibility of differences between *explanations* in biological science and physical science:

> Despite the undeniable successes of physicochemical explanations in the study of living things, biologists of unquestioned competence continue to regard such explanations as not entirely adequate for the subject matter of biology. … Some of them nevertheless maintain that the mode of analysis required for understanding living phenomena is fundamentally different from that which obtains in the physical sciences.[9]

Nagel identified two reasons why biological inquiry and explanation might differ from that found in physical science: "One is the dominant place occupied by teleological explanations in biological inquiry. The other is the use of conceptual tools uniquely appropriate to the study of systems whose total behavior is not the resultant of the activities of independent components."[10] The latter reason can be rephrased in terms of (the failure of) part-whole explanations. In his extensive discussion of these reasons, Nagel made a salient observation about two different

8 Simpson (1964, pp. 106-107)
9 Nagel (1961, p. 398)
10 Nagel (*Ibid.*, p. 401)

modes of analysis that correspond to the distinction between structure and function in biology.

> The contrast between structure and function is evidently a contrast between the *spatial* organization of anatomically distinguishable parts of an organ and the *temporal* (or spatio-temporal) organization of changes in those parts. What is investigated under each term of the contrasting pair is a mode of organization or a type of order. In the one case the organization is primarily if not exclusively a spatial one, and the object of the investigation is to ascertain the spatial distribution of organic parts and the modes of their linkage. In the other case the organization has a temporal dimension, and the aim of the inquiry is to discover sequential and simultaneous orders of change in the spatially ordered and linked parts of organic bodies.[11]

This observation and many others demonstrate Nagel's acute awareness of the subtle relations between explanations in biology and explanations in physics. Although his conclusion was cautiously optimistic about the future possibility of offering complete physicochemical explanations of biological phenomena, Nagel recognized that the resistance to explanatory reduction by organismal biologists was motivated by an important point: "the stress they place on the hierarchical organization of living things and on the mutual dependence of organic parts is not a misplaced one."[12]

Both Mayr and Nagel spoke in very general terms about biology and physics, their explanatory modes, and the success or failure of reductionism. Many biologists and philosophers also have worried that importing models of reasoning from the physical sciences obscures our understanding of reasoning in the life sciences in more circumscribed domains. For example, teleological or functional explanations continue to be an area of debate in philosophy of biology but the topic is absent from philosophy of physics. In this paper we follow a more circumscribed strategy by combining three elements present in the dialectic between Mayr and Nagel: explanatory reduction, part-whole relations, and the temporal dimension of organization. Biology and physics tend to incorporate different models of temporality in part-whole reductive explanations, which partially validates the worry that modes of analysis in physics cannot be imported directly into philosophical analyses of inquiry and explanation in biology. After documenting possible differences in temporal aspects of part-whole reductive explanations, we argue that these result from differential emphases on compositional and causal facets of reductive explanations. Keeping these two facets distinct facilitates the identification of two further aspects of reductive explanation: intrinsicality and fundamentality. The result is an account that provides resources for discriminating between different types of reductive explanation, and suggests a new approach to comprehending similarities and differences in the explanatory reasoning found in biology

11 Nagel (*Ibid.*, p. 426)
12 Nagel (*Ibid.*, p. 444)

and physics, without having to decide the general question of whether and in what sense biology is autonomous from physics.

2. Temporality in Part-Whole Reductive Explanations

2.1 Part-Whole Reductive Explanations

A variety of authors have conceptualized reductionism in terms of the relationship between parts and wholes.[13] Nagel also devoted attention to part-whole relations,[14] but his analysis was overshadowed in subsequent developments of theory reduction by philosophers (akin to Mayr overlooking aspects of Nagel's views on explanatory reduction). Part-whole relations play both methodological and explanatory roles in biology and are distinct from identity-reduction, which focuses on the relation of two kinds of properties within the same system rather than explaining properties of a compound system in terms of its parts. Explaining the property of a whole in terms of the properties of its parts is a distinct explanatory question from explaining a property of the whole in terms of another property of the whole by identity.

Consider a case where the property of a whole (the temperature of an ideal gas) is explained in terms of the combined properties of the parts (kinetic energies of the molecules). One question we can ask is whether it is possible to explain why an ideal gas with a certain mean kinetic energy has a specific temperature. This involves two kinds of properties of the same system—the specific mean kinetic energy and temperature of the ideal gas at a time. A correspondence rule may link the mean kinetic energy (a property of the gas as a whole) to the temperature (a property of the gas as a whole). A second question is whether it is possible to explain why an ideal gas has a specific mean kinetic energy on the basis of the behavior of its constituent molecules. This involves the behavior of the components of the ideal gas and its behavior as a whole: how do the individual kinetic energies add up to the kinetic energy of the whole? The relation is between parts and wholes, not two properties within the same system.

2.2 Temporality

In Nagel's contrast between spatial and temporal modes of organization, the temporal mode is characterized by its concentration on "sequential and simultaneous orders of change in the spatially ordered and linked parts of organic bodies." These orders of change are described frequently in causal terms, such as the parts of an organism exhibiting causal interactions conditioned by spatial proximity. All causal explanations involve some element of temporal duration. If the aim is to

13 e.g., Bechtel and Richardson (1993); Sarkar (1998); Wimsatt (1976)

14 Nagel (*Ibid.*, pp. 380-397)

explain the increased rhythmic contraction of the heart by the cellular mechanisms that process adrenergic hormones, then the explanation requires (at least implicitly) an earlier time at which hormones are interacting with cellular receptors and a later time when the altered rhythmic contraction obtains. Temporal duration may be operationalized in different ways depending on the explanatory goals in view, such as with absolute chronology (minutes or hours) or event sequences or stages (Fertilization → Cleavage → Blastula → Gastrula).

We can make the role of time more precise by characterizing temporality for part-whole reductive explanations in terms of a property of a whole at t^* being explained by properties of its parts at an earlier time t. If a *temporal* relation is one in which a property or state at t is related to another property or state at t^*, then a *causal* relation is one in which a property or state at t determines or influences another property or state at t^*; the state or properties of the parts and their interactions at t (or $t_1, \ldots, t_n < t^*$) bring about a change in the state or properties of the compound at time t^*. Many concepts commonly invoked in philosophical discussions of reduction are atemporal and non-causal.[15]

Part-whole explanations in physics are often atemporal.[16] For the behavior of a physical system, part-whole reductions can pertain to either its states or its temporal evolution (i.e., dynamics). A part-whole reduction of a *state* explains the state of a compound system at a time on the basis of the states of its parts at the same time. For example, we might explain the determinate energy value E (macrostate) of a compound system (e.g., an ideal gas) by appeal to the determinate energy values of its constituents (e_1 to e_n; the states of the parts). The explanation relies on the states of the parts (particular facts) and a law of composition delineating how these states contribute to the state of the compound. If we assume that interactions can be neglected, the kinetic energy values simply add up. The explanation is reductive, because it only uses states of the parts and a law of composition in the *explanans*, but it is also atemporal.[17]

A second situation is the part-whole reduction of the *dynamics* of a physical system; the temporal evolution of a compound system can be explained in terms of the dynamics of its parts. For instance, the Hamilton operator for a compound system can be analyzed in terms of those for the parts along with interaction terms and a law of composition. The relation between the Hamilton operator

15 Supervenience is one example: 'Mental properties supervene on physical properties, in that necessarily, for any mental property M, if any thing has M at time t, there exists a physical base (or subvenient) property P at t, and necessarily anything that has P at a time has M at that time' (Kim 1998, 9).

16 Hüttemann (2005)

17 The situation is more complex for systems of the order of 10^{23} degrees of freedom, as in statistical physics. Averaging procedures such as 'coarse graining' are necessary to explain the behaviour of compound systems. These procedures often involve averaging over time so that the relations in question are no longer atemporal in the sense defined here.

for the compound and those for the parts (i.e., a part-whole relation) is reductive but *a*temporal because the Hamilton operator of the compound at time t^* is not calculated in terms of the Hamilton operators of the parts and their interactions at *another* time t. (Typically, the Hamilton operators of the parts and the compound are time-independent.) A part-whole reduction of the dynamics of a compound is often integrated into a causal explanation. In classical dynamics, if we want to explain why the state of the solar system Z develops over time into state Z', we can appeal to the dynamics of its parts and their interactions. The two dimensions of the explanation are separable: (a) the non-reductive, temporal dimension; and, (b) the atemporal, part-whole reductive dimension.

Part-whole reductions (and explanations more generally) in biological science are often *temporal*. Properties of a whole at t^* are explained in terms of properties of parts at an earlier time t (or $t_1, \ldots, t_n < t^*$); the behavior of the parts at earlier times *causes* the compound to have a behavior or property at a later time t^*. Temporal or causal part-whole reductions can be described differently depending on how the behavior of the parts is described. When laws play a role, causal part-whole reductions involve the laws and initial conditions that pertain to parts of a whole at t explaining the behavior of the whole at t^*. Causal part-whole reductions also can be described in terms of causal powers: the behavior of a whole at t^* is explained in terms of the causal powers of its parts at t. Causal part-whole explanations are reductive because they constrain the *explanans* to only laws or causal powers of the parts (and sometimes compositional rules).

Examples of causal part-whole reductions abound in biology, such as explaining muscle tissue activity (behavior of a system) at a later time t^* by appealing to the properties of muscle cells (the parts) composing the tissue (the whole), which contain special motor proteins that contract by molecular ratcheting, at an earlier time t. Temporal part-whole reductions explain the state of a compound or whole (muscle contraction) in terms of states of the parts at earlier times (myosin ratcheting) and must, for several reasons, be distinguished from atemporal part-whole reductions, which explain the behavior of a compound in terms of the behavior of the parts at the same time.

3. Composition, Causation, and the Difference Time Makes

3.1 Composition and Causation

What explains the difference in how temporality enters into part-whole reductive explanations in biology and physics? The answer lies in the distinction between composition and causation, both of which are key facets of reductive explanation. Composition refers to higher-level entities being constituted by, realized by, or nothing but lower-level entities. The heart is composed of myocardial cells and other cellular entities. A myosin filament found in myocardial cells is composed

of myosin proteins, which are in turn composed of amino acid residues. Causation refers to higher-level entities being caused, brought about, or determined by lower-level entities. The heart's rhythmic beating is caused by the contraction of its myocardial cells. Myocardial cells contract because myosin proteins ratchet along another set of proteins. The appropriate folding of a particular myosin protein is determined causally by its amino acid sequence.

Atemporal part-whole reductions, which correspond to what one typically finds in physical science, focus on *composition*; i.e., the relation of the higher level to the lower level is about constitution or realization. When there is a temporal element of the explanation, such as when explaining dynamics, the element of causation only enters as a relation between entities at the same level (e.g., the evolution of the state of the parts from an earlier to a later time). Temporal part-whole reductions, which are characteristic of reductive explanatory practices in biology, combine the compositional element with that of *causation*; they contain a mixture of compositional and causal claims.[18] In the part-whole reductive explanation for a physical system described above, the temporal (causal) and the reductive (part-whole) dimensions can be separated neatly in contrast to the intertwined compositional and causal claims found in biology.

We agree with Nagel that this difference does not establish "the absolute autonomy of biology or the inherent impossibility of physicochemical explanations of vital phenomena."[19] Therefore, we leave open the possibility that temporal part-whole reductions from biology might be explicated in terms of atemporal part-whole reductions in physics at some point in the future (according to standards of the scientific community). But the difference in explanatory practice motivates distinguishing the two facets. In particular, the way temporality enters into part-whole reductive explanations directs our attention to two further aspects of reductive reasoning that reveal more resources for characterizing differences between modes of analysis in biology and physics.

3.2 Intrinsicality and Fundamentality

Part-whole spatial relations presume that parts are contained within or intrinsic to the whole. Thus, *intrinsicality* seems to be a precondition for part-whole reductive explanations, which relates directly to composition. Claims about reducing a higher-level entity to its component parts involves an individuation of the higher-level entity such that its components can be distinguished from other entities (a surrounding context or environment). If the goal is to reductively explain a cell in terms of its component parts, then the intrinsic/extrinsic boundary is the cell membrane. If the goal is to reduce the heart to its component parts then the boundaries

18 Craver and Bechtel (2007)

19 Nagel (*Ibid.*, p. 444)

of the organ demarcate intrinsic from extrinsic. What counts as intrinsic is relative to local explanatory aims.[20]

Parts also are considered more fundamental than the whole; they are the 'lower level' that reductively explains the 'higher level' system properties of the whole (*fundamentality*[21]). In many cases, only a subset of fundamental level properties will count as explanatory (e.g., biochemical moieties but not spatial location). As with intrinsicality, these qualifications are spelled out locally in the context of explanation. But they also introduce qualifications about intrinsicality. A qualified fundamentality might identify biochemical properties as explanatory of cell properties, which implies that biochemical properties outside of the cell (i.e., extrinsic) are fundamental and that other intrinsic properties of cellular components (e.g., location) are not explanatory.

Intrinsicality and fundamentality are constraints on part-whole reductive explanations. Thus, compositional part-whole reductions can fail as *reductive* explanations either because intrinsicality is violated (for a particular part-whole decomposition) or fundamentality is violated (in the broad sense or for some restricted set of properties). But it seems impossible for these to be violated independently because, at any particular time, the parts are intrinsic *and* fundamental to the whole. If biochemical properties of a cell's 'intrinsic' parts are considered 'fundamental', then they either explain the cell level properties or they do not. This is where the significance of temporality emerges because it allows for the decoupling of intrinsicality and fundamentality in part-whole reductive explanations. If part-whole relations are considered temporally (as expected for biology), then it is possible for intrinsicality to be violated because a part of a whole at time t may no longer be a part of the whole at t^*.

Heart organogenesis is one empirical case where we can observe the importance of temporality for part-whole reductive explanations. Blood cells coursing through nascent atrial chambers during ontogeny meet the intrinsicality condition (spatially) at time t even though at t^* they have passed out of the heart. Fluid flow is a key factor in the proper shaping of the heart during embryogenesis;[22] blood cell components at t have a causal effect on the shape of the heart at t^* when these cells are no longer intrinsic. This is still a kind of reductive explanation because properties of entities at the qualified fundamental level of cells account for changes in the properties of entities at the non-fundamental level (organs). Another common example is programmed cell death (apoptosis). Cellular components at time t

20 Researchers are explicit about this; e.g., "it is often beneficial to separate contributions arising from fluctuations that are inherent to the system of interest (intrinsic noise) from those arising from variability in factors that are considered to be external (extrinsic noise). … The definition of intrinsic noise is problem-dependent, and varies from one context to another" (Kærn et al. 2005, p. 456).

21 see Sarkar (1998, ch. 3)

22 Hove et al. (2003)

can bring about a change in the properties of a whole at *t** even though the cellular components are literally absent at *t** due to apoptosis.

Two other kinds of possibilities bear mentioning. First, adrenergic hormones secreted in the circulatory system can modulate heart rhythm. These hormones are extrinsic to the heart, even though they meet the fundamentality condition, and exert an effect through a temporally extended process. Second, different organs can interact directly (through physical contact) with the heart to bring about changes in its morphology during embryogenesis. These other organs are not only extrinsic but also non-fundamental because they are at the 'same' non-fundamental level as the heart.

We can summarize these possibilities in a table of reductive explanatory forms (Table 1). This displays the heterogeneous ways that part-whole explanations can succeed or fail as *reductive* explanations.[23]

Table 1

Forms of Explanation	*Temporality*	*Intrinsicality*	*Fundamentality*
I	N	Y	Y
II	N	N	N
III	Y	Y	Y
IV	Y	N	Y
V	Y	N	N

The first two forms (I and II) correspond to purely compositional relations (I = success; II = failure). Form III includes the possibility of successfully extending compositional relations through time and also encompasses explanations that appeal to fundamental, intrinsic causes at *t* bringing about an effect in the non-fundamental level whole at *t** while ignoring compositional relations that obtain at other times between *t* and *t** (such as blood flow in the heart or apoptosis). Form IV concerns an extrinsic entity from a fundamental level explaining a system property (e.g., the adrenergic hormone case). Although it is reductive because it appeals to properties at the fundamental level, it fails in the sense of a whole being causally explained by its parts. Form V corresponds to the possibility of one organ interacting with another organ—the entity is extrinsic and also resides at the same level of non-fundamentality. Here there is a failure of the parts to explain the whole *and* a failure of fundamental level properties to explain non-fundamental level properties. Once part-whole relations are temporally indexed, intrinsicality and fundamentality take on independent significance in reductive explanations.[24]

23 Table 1 suppresses the diversity that obtains as a result of individually characterizing intrinsicality and fundamentality for a particular reductive explanation.

24 The table omits two kinds of possibilities ruled out in discussion: (a) atemporal part-whole reductions that violate either intrinsicality or fundamentality alone (two forms of explanation), and, (b) fundamentality failing when intrinsicality holds, because if a

Table 1 shows the difference that time makes. Without it, there is only a question of whether the part-whole explanatory reduction succeeds (I) or fails (II). But the successes and failures of temporal part-whole explanatory reductions in biology often involve forms III, IV, and V. Now it is clear why the export of atemporal models of reductive explanation from philosophical reflection on physical science could hinder our analyses of reductive explanations in biology. In the part-whole reductive explanation for a physical system described above, the temporal (causal) and the reductive (part-whole) dimensions can be separated in contrast to the mixed compositional and causal claims found in biology or situations where the compound does not exist at the earlier time (e.g., during embryological development). Furthermore, physics typically treats isolated systems and, as a consequence, it is often assumed that parts are *not* added or lost from the compound (as happens during cell division or apoptosis); i.e., it is assumed that compositional relations remain constant. Our analysis makes explicit why some physics-derived models mischaracterize part-whole reductive explanations found in biological reasoning; i.e., in what respect explanations in biology and physics can be different.

In addition to this fault line between modes of analysis in biology and physics, our account generates a new perspective on the 'context' objection to reductive explanation.[25] The context objection claims that a reduction can be blocked because of an ineliminable appeal to contextual factors. A standard rejoinder is to pursue a reduction of this context. The inclusion of temporality gave us two different ways a reductive explanation can fail as a *reductive* explanation (IV – intrinsicality fails; V – fundamentality and intrinsicality fail). This provides the basis for a more nuanced reading of the context objection. The reductionist rejoinder maintains the fundamentality condition even when intrinsicality is violated; it preserves fundamentality (a reductive explanation in one aspect) at the expense of intrinsicality, and thus the context objection retains some validity. If an attempt is made to recover intrinsicality by redrawing system boundaries (e.g., treat the body cavity as the system so that the adrenergic hormone is a part of this new whole), then we have changed what counts as whole and parts, and thereby what counts as intrinsic. Whether a part-whole reductive explanation succeeds or fails with respect to intrinsicality depends on how 'wholes' are individuated, which means redrawing boundaries constitutes a change of the original question about whether a whole can be explained reductively by its parts. The failure of a reductive explanation for the aspect of intrinsicality may be an empirical indicator that redrawing boundaries is warranted epistemologically. But even if a reductive explanation succeeds as a consequence of new individuation criteria, this is compatible with the claim that the behavior of interest cannot be explained in terms of its intrinsic features under the original individuation criteria. Thus, the success or failure of

feature is intrinsic then in order to be contained within a whole it must be instantiated at a more fundamental level than the whole.

25 cf. Delehanty (2005)

a reductive explanation is not an all or nothing phenomenon. For any part-whole reductive explanation, we must not only inquire whether each of these aspects is applicable, but also characterize the details involved in order to evaluate whether there is success or failure *of one kind or another*.

4. Examples: Part-Whole Reductive Explanations in Biology and Physics

4.1 Ideal Crystal

One example of a reductive part-whole explanation in physics is the classical treatment of the ideal crystal. It is a reductive explanation of the dynamics (temporal evolution) of a compound system in terms of the dynamics of the parts and their interactions. According to standard treatments, the electrons and ions that constitute the crystal can be considered separately (*adiabatic approximation*). The regular structure of the crystal is generated by the ions. Within the so-called harmonic approximation they are, however, not supposed to sit motionless at their lattice-sites. According to the model the ions perform oscillations around the sites of the lattice, which are described as the mean equilibrium positions of the ions. These oscillations are considered small in comparison with the inter-ionic spacing, which means that only nearest-neighbor interactions are relevant. Furthermore, it is supposed that the potential between nearest neighbors is harmonic.[26] On the basis of these assumptions we can specify the classical Hamilton function of the ideal crystal. The Hamilton function is constructed in terms of the dynamics of the constituents, which are understood as isolated (kinetic energy terms), and their interactions (potential energy). These contributions are added together according to a law of composition

$$H = \sum_i E^i_{kin} + (1/2) \sum_{ij} U_{ij} q_i q_j \qquad (4.1.1)$$

where $E^i_{kin} = p^2_i / 2m$ is the kinetic energy of the parts, and $U_{ij} = \partial^2 / \partial q_i \partial q_j \, U(q_1 \ldots q_{3N})$ describes the interactions between the parts. On the basis of the Hamilton function we can determine the thermal density of the crystal, which is given by

$$u = 1/V \, (\int d\Gamma \, exp \, \{-\beta H\} H) / (\int d\Gamma \, exp \, \{-\beta H\}) \qquad (4.1.2)$$

in which $d\Gamma$ stands for the volume element in crystal phase space and $\beta = 1/k_B T$ where k_B is the Boltzmann-constant and T the temperature. The thermal density of the crystal permits us to calculate the behavior of the compound system, including measurable thermodynamic properties such as the specific heat c_v:

$$c_v = (\partial / \partial T) \, u \qquad (4.1.3)$$

Classically, the specific heat in a crystal is independent of its temperature.

26 Ashcroft and Mermin (1976, pp. 422-427)

Within the conceptual framework developed above (Sections 2 and 3), this example of a reductive physical explanation has two important elements. First, the explanation is *atemporal* because the Hamiltonian of the compound for a certain time t does not depend on the Hamiltonians for the parts at another time t^* (this obtains trivially because the Hamiltonian is time-independent). Second, the explanation relies only on features that are intrinsic with respect to the compound, which is typical for many areas of physics that focus on isolated systems. Furthermore, the properties appealed to in the explanation are those of the parts alone, and in this sense the aspect of fundamentality is met. The explanation of the crystal's behavior provides an example of a reductive part-whole explanation that corresponds to Form I in Table 1.

4.2. Quantum-entanglement

Not all reductive part-whole explanations of states (as opposed to the dynamics) in physical science correspond to what we see in the case of the ideal crystal. Consider the spin states of a compound that consists of two non-identical particles. Normalized vectors in two-dimensional Hilbert spaces represent the spin states of the separate particles: H_1 and H_2. In order to construct a Hilbert-space for the compound system we need a law of composition; e.g., the possible spin states of the compound system are all those states that can be represented as (normalized) vectors in the tensor product of H_1 and H_2 : $H_s = H_1 \otimes H_2$.

If we take the eigenvectors in the spin z-direction as the basis for H_1 and H_2 (H_1: $|\psi^{z\text{-}up}{}_1>$ and $|\psi^{z\text{-}down}{}_1>$; H_2: $|\psi^{z\text{-}up}{}_2>$ and $|\psi^{z\text{-}down}{}_2>$), then we find all of the following among the possible states of the compound system:

(4.2.1) $|\psi^{z\text{-}up}{}_1> \otimes |\psi^{z\text{-}down}{}_2>$

(4.2.2) $|\psi^{z\text{-}down}{}_1> \otimes |\psi^{z\text{-}up}{}_2>$

(4.2.3) $1/\sqrt{2}\,|\psi^{z\text{-}up}{}_1> \otimes |\psi^{z\text{-}down}{}_2> - 1/\sqrt{2}\,|\psi^{z\text{-}down}{}_1> \otimes |\psi^{z\text{-}up}{}_2>$

What is essential is that equation 4.2.3 cannot be written as a simple tensor product of vectors H_1 and H_2; it can only be written as a superposition of such tensor-products. The fact that the compound is in a determinate state cannot be explained in terms of the determinate states the constituents occupy. This is because there are states, such as those described in equation 4.2.3, which do not allow the attribution of pure states to the parts of the compound. A part-whole explanation of the state of the compound thus fails.

This is a failure of reductive explanation because a part whole-explanation of the state is not merely difficult to formulate but impossible to achieve. Quantum mechanics contains states of compound systems that do not allow for the attribution of pure states to the parts. The impossibility of attaining these kinds of reductive explanations is implied by the formalism of quantum mechanics. Reductive explanations that correspond to Form I in Table 1 are not obtainable. The case of

spin-states can be classified as an instance of Form II. In both of the cases from physical science, fundamentality is understood in terms of properties or states of the parts; intrinsicality is assumed. Thus, fundamentality and intrinsicality stand or fall together. Section 3 described how these two aspects could be decoupled with the addition of temporality, and biological science is a natural place to look for this explanatory pattern.

4.3 Protein Folding

Biologists have long recognized that part-whole reductive explanations are relevant in the context of explaining how protein folding occurs: "The protein folding problem … represents an unusually concrete and limited case of the whole problem of reductionism. … understanding the rules of [folding] would teach us worthwhile lessons about … exactly how an organic whole becomes so much more than a sum of its parts."[27] Philosophers have also observed that it represents a key locus for evaluating part-whole explanatory reduction in molecular biology.[28] Proteins are composed of amino acid components ('residues') that are linked by covalent peptide bonds into a chain ('polypeptide'). This linear chain is produced from a process termed 'translation'; specific cellular constituents (ribosomes, themselves proteins) translate a linear stretch of RNA with a triplet code of nucleotides (e.g., AAG) into amino acid residues for a linear polypeptide (e.g., AAG = lysine). Nearly all proteins adopt a three-dimensional structure in order to be functional, which is understood in terms of interactions among its amino acid residues (e.g., hydrophobic residues avoid interaction with surrounding water by segregating to internal regions). Addressing the protein folding problem requires explaining how this conformation is achieved for polypeptides subsequent to translation from RNA in the cellular context.[29]

The linear sequence hypothesis holds that the three-dimensional folding of a protein results from the properties of the amino acid residues in the polypeptide and their chemical interactions alone—the whole is a 'sum' of the interaction of its parts. Although there is an ambiguity in the linear sequence hypothesis between (a) inferring or predicting the three dimensional structure of a protein from its linear sequence of amino acids, and (b) explaining the outcome of three dimensional structure by appeal to the kinetic, thermodynamic, and structural processes in the cell,[30] we focus only on the latter construal because it concerns part-whole explanatory reduction. Whether a protein folds only as a consequence of its amino acid residues is a causal question involving reductive explanations of wholes in terms of parts.

27 Richardson (1982, p. 1)

28 Sarkar (1998, p. 169)

29 Some folded proteins aggregate further in order to be functional (e.g. hemoglobin is a tetramer).

30 cf. Freedman (1999)

Evidence in favor of the linear sequence hypothesis was derived initially from experiments on the denaturation and refolding of ribonuclease proteins *in vitro*.[31] Ribonucleases subjected to denaturing conditions were able to refold rapidly into the proper configuration. Correct refolding seemingly occurred as a function of the linear sequence of amino acid residues composing the polypeptide. But the folding took an hour or longer rather than several minutes or less without an enzyme from the endoplasmic reticulum (a cellular organelle where much translation occurs). Many denatured proteins do not refold as cleanly as those studied by Anfinsen's group[32] and the process requires the activity of chaperone proteins that guide folding during and after polypeptide synthesis[33]: "Proteins need the assistance of molecular chaperones and folding enzymes to reach their native structure efficiently".[34]

Molecular chaperones must provide oversight during folding because the cellular environment is crowded.[35] Distinct functional groups of chaperones monitor and facilitate protein folding during *de novo* synthesis, quality control, and the response to stress.[36] Multiple amino acid residue interactions between an already functional, folded protein (the chaperone) and the as-of-yet folded polypeptide underlie the process of correct folding.[37] Even when mutations are introduced that lead to altered amino acid components in a polypeptide, which should prevent correct folding, proper folding can be induced by the overproduction of molecular chaperones.[38]

One way the linear sequence hypothesis might fail is that the ordering of the amino acid residues may be insufficient to explain the three dimensional conformation of the folded protein, assuming the laws of macromolecular physics.[39] Two separate constraints operate in the temporal part-whole reductive explanation offered by the linear sequence hypothesis. First, only properties of the parts are required to explain protein folding (i.e., intrinsicality). A property is intrinsic to the linear polypeptide if it is a property of one of its amino acid components or their interactions (intrinsic relational properties). Contextual or extrinsic causal factors are not supposed to play an essential role or contribute to correct folding, such as physico-chemical components (e.g., H_2O), other proteins (e.g., chaperones), or nucleic acids (e.g., RNA). Second, the amino acids ('parts'), as well as macromolecular laws that describe their interactions, are available to explain protein folding (fundamentality). System properties due to a complex three-dimensional

31 Anfinsen (1973)
32 Clark (2004)
33 Feder and Hofmann (1999); Frydman (2001)
34 Liscalijet et al. (2005, p. 78)
35 Ellis (2001); Homouz et al. (2008); Liscalijet et al. (2005)
36 Albanese et al. (2006); Ellis (1998); McClellan et al. (2005); Tang et al. (2006)
37 Tang et al. (2008)
38 Maisnier-Patin et al. (2005)
39 Sarkar (1998, pp. 169-170)

structure that are absent from the linear polypeptide are not fundamental. The tertiary structure of a three-dimensional protein 'whole' is explained by the interaction of its component parts at earlier times; the amino acid residues interact causally to bring about the state of the whole ('correctly folded').

Molecular chaperones are non-intrinsic, non-fundamental causal factors that make specific and necessary contributions to folding (not just as appropriate environmental background): "The manner in which a newly synthesized chain of amino acids transforms itself into a perfectly folded protein depends both on the intrinsic properties of the amino-acid sequence and on multiple contributing influences from the crowded cellular milieu."[40] The intrinsic properties of the linear polypeptide arising from its amino acid residue parts are not sufficient to explain the manifestation of protein folding. The temporally extended process of folding not only requires appropriate environmental conditions but also the contribution of extrinsic chaperones; i.e., there is a failure with respect to the aspect of intrinsicality. Additionally, the causal contribution of chaperones in protein folding results from three-dimensional structure, a kind of property the amino acid parts lack. Thus, the best explanation of protein folding also involves a failure with respect to the aspect of fundamentality (Form V in Table 1). Systems with properties due to complex three-dimensional structure (folded proteins), rather than systems that lack it, are necessary to produce the native conformations of proteins *in vivo*—the parts alone in combination with the macromolecular laws of composition are not enough. Temporal part-whole reduction fails with respect to both aspects as a *reductive* explanation.[41]

Could a 'reductionist' adopt the rebuttal to the context objection here ('just reduce the context also')? Chaperones are composed of parts and therefore we can 'reduce' the operation of an extrinsic chaperone protein whole to its parts. This is akin to the strategy of preserving a reduction by 'extending the mechanism' backwards in time.[42] But chaperone proteins require other chaperones for their own proper folding, so the attempt to reduce the extrinsic chaperone (or extend the mechanism) leads to a type of explanatory regress. According to the individuation schemes adopted by scientists, extrinsic, non-fundamental wholes (folded proteins—chaperones) are required for the proper folding of another whole (folded protein). A related objection is to suggest a new individuation scheme: the cell as a 'larger' whole contains the protein *and* the crowded cellular milieu, thereby mak-

40 Dobson (2003, p. 884). "There is a need for molecular chaperones because the intrinsic properties of proteins assure that incorrect interactions are possible" (van der Vies et al. 1993, p. 73).

41 This claim is relative to the individuation and decomposition of the system offered by scientists, and pertains to the process of *bringing about* the three-dimensional protein-structure. Whether the pertinent causal powers of the molecular chaperones are truly novel *vis-à-vis* the causal powers of its parts concerns constitutional reductionism in an atemporal sense rather than causal part-whole reductive explanation.

42 Delehanty (2005)

ing the molecular chaperones intrinsic and fundamental. But even if a causal part-whole explanation of the behavior of the *cell* is feasible in terms of *its* intrinsic parts, this would change the question of what parts and wholes are being reduced. The folding of a *protein* still cannot be explained solely in terms of *its* amino acid parts; both intrinsicality and fundamentality are violated because the features required to explain folding are extrinsic and not located in the fundamental realm. Shifting to a larger whole simply changes the *explanandum*.

5. Conclusion

The examples canvassed in Section 4 illustrate that part-whole explanations in biology often work differently from those in physics. This is due to the fact that part-whole explanations in physics primarily focus on compositional relations, whereas biological part-whole explanations focus on whether the behavior of the parts at an earlier time *t* cause the behavior of the compound at *t** (Sections 2 and 3). As a consequence there are different ways for reductive explanations to succeed or fail. Because fundamentality is often spelled out in terms of the properties of the parts alone (i.e., intrinsicality) in physical science explanations, a failure of reductive explanation is a failure with respect to both aspects. Furthermore, since physics typically deals with isolated systems, there is no way to make an explanatory appeal to extrinsic features. Thus, for spin-states of a compound, the failure of reductive explanation is also a failure of explanation, and has often been classified as an emergent phenomenon.[43]

In contrast, the biological case illustrates how a reductive explanation can fail in different ways (intrinsicality, fundamentality, or both) without automatically failing as an explanation. Protein folding has been explained, just not only in terms of the fundamental, intrinisic properties of amino acid residues composing the polypeptide. Importing physical science conceptions of reductive part-whole explanations into biology may obscure this important point. So Mayr was correct that there are genuine dangers for such an import. However, we do not agree with his more sweeping claims concerning explanatory reduction. By developing a conceptual framework for reductive part-whole explanations it is easier to see in what respects such explanations in physics and biology differ. The point of departure for this framework was Nagel's observation of the distinctness of temporal modes of organization in biological systems (Section 1). Explicitly incorporating temporality provides the basis for decoupling fundamentality and intrinsicality (Section 2), which yields multiple combinations of success and failure for reductive explanations that emerge from distinguishing composition and causation (Section 3; Table 1). It also offers a more precise foundation for delineating differences *and* similar-

43 Humphreys (1997)

ities between the explanatory approaches of biology and physics. The result is an increased comprehension of how reductive explanations operate in the sciences.

References

Veronique Albanese, Alice Yen-Wen Yam, Joshua Baughman, Charles Parnot, and Judith Frydman, "Systems analyses reveal two chaperone networks with distinct functions in eukaryotic cells", in: *Cell* 124, 2006, pp. 75-88.

Christian B. Anfinsen, "Principles that govern the folding of protein chains", in: *Science* 181, 1973, pp. 223-230.

Neil W. Ashcroft and N. David Mermin, *Solid State Physics*. Philadelphia, Cengage Learning Services 1976.

John Beatty, "The proximate/ultimate distinction in the multiple careers of Ernst Mayr", in: *Biology and Philosophy* 9, 1994, pp. 333-356.

William Bechtel and Robert C. Richardson, *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton: Princeton University Press 1993.

Patricia L. Clark, "Protein folding in the cell: reshaping the folding funnel", in: *Trends in Biochemical Sciences* 29, 2004, pp. 527-534.

Carl. F. Craver and William Bechtel, "Top-down causation without top-down causes", in: *Biology and Philosophy* 22, 2007, pp. 547-563.

Megan Delehanty, "Emergent properties and the context objection to reduction", in: *Biology and Philosophy* 20, 2005, pp. 715-734.

Christopher M. Dobson, "Protein folding and misfolding", in: *Nature* 426, 2003, pp. 884-890.

R. John Ellis, "Steric chaperones", in: *Trends in Biochemical Sciences* 23, 1998, pp. 43-45.

R. John Ellis, "Macromolecular crowding: obvious but underappreciated", in: *Trends in Biochemical Sciences* 26, 2001, pp. 597-604.

Martin E. Feder and Gretchen E. Hofmann, "Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology", in: *Annual Review of Physiology* 61, 1999, pp. 243-282.

Robert B. Freedman, "Protein folding in the cell", in: Thomas E. Creighton (Ed.), *Protein Folding*. New York: W.H. Freeman and Company 1999, pp. 455-539.

Judith Frydman, "Folding of newly translated proteins in vivo: the role of molecular chaperones", in: *Annual Review of Biochemistry* 70, 2001, pp. 603-647.

Dirar Homouz, Michael Perham, Antonius Samiotakis, Margaret S. Cheung, and Pernilla Wittung-Stafshede, "Crowded, cell-like environment induces shape changes in aspherical protein", in: *Proceedings of the National Academy of Sciences of the United States of America* 105, 2008, pp. 11754-11759.

Jay R. Hove, Reinhard W. Köster, Arian S. Forouhar, Gabriel Acevedo-Bolton, Scott E. Fraser, and Morteza Gharib, "Intracardiac fluid forces are an essential epigenetic factor for embryonic cardiogenesis", in: *Nature* 421, 2003, pp. 172-177.

Paul Humphreys, "How properties emerge", in: *Philosophy of Science* 64, 1997, pp. 1-17.

Andreas Hüttemann, "Explanation, emergence, and quantum entanglement", in: *Philosophy of Science* 72, 2005, pp. 114-127.

Mads Kærn, Timothy Elston, William Blake, and James Collins, "Stochasticity in gene expression: from theories to phenotypes", in: *Nature Reviews Genetics* 6, 2005, pp. 451-464.

Jaegwon Kim, *Mind in a Physical World*. Cambridge, MA: MIT Press 1998.

I. Marije Liscalijet, Bertrand Kleizen, and Ineke Braakmen, "Studying protein folding in vivo", in: Johannes Buchner and Thomas Kiefhaber (Eds.), *Protein Folding Handbook. Part II*. Weinheim: WILEY-VCH Verlag 2005, pp. 73-104.

Sophie Maisnier-Patin, John R. Roth, Asa Fredriksson, Thomas Nystrom, Otto G. Berg, and Dan I. Andersson, "Genomic buffering mitigates the effects of deleterious mutations in bacteria", in: *Nature Genetics* 37, 2005, pp. 1376-1379.

Ernst Mayr, "Cause and effect in biology", in: *Science* 134, 1961, pp. 1501-1506.

Ernst Mayr, *Toward a New Philosophy of Biology: Observations of an Evolutionist*. Cambridge, MA: Harvard University Press 1988.

Ernst Mayr, *What Makes Biology Unique? Considerations on the Autonomy of a Scientific Discipline*. New York: Cambridge University Press 2004.

Amie J. McClellan, Melissa D. Scott, and Judith Frydman, "Folding and quality control of the VHL tumor suppressor proceed through distinct chaperone pathways", in: *Cell* 121, 2005, pp. 739-748.

Ernest Nagel, *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace & World, Inc 1961.

J. S. Richardson, "What do the folds in proteins look like?", in Donald B. Wetlaufer (Ed.), *The Protein Folding Problem.* Boulder, CO: Westview Press 1982, pp. 1-28.

Sahotra Sarkar, *Genetics and Reductionism*. Cambridge: Cambridge University Press 1998.

George G. Simpson, *This View of Life: The World of an Evolutionist*. New York: Harcourt, Brace & World 1964.

Yun-Chi Tang, Hung-Chun Chang, Annette Roeben, Dirk Wischnewski, Nadine Wischnewski, Michael J. Kerner, F Ulrich Hartl, and Manajit Hayer-Hartl, "Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein", in: *Cell* 125, 2006, pp. 903-914.

Yun-Chi Tang, Hung-Chun Chang, Kausik Chakraborty, F. Ulrich Hartl, and Manajit Hayer-Hartl, "Essential role of the chaperonin folding compartment in vivo", in: *EMBO Journal* 27, 2008, pp. 1458-1468.

Saskia M. van der Vies, Anthony Gatenby, Paul V. Viitanen, and George H. Lorimer, "Molecular chaperones in and their role in protein assembly", in: Jeffrey L. Cleland (Ed.), *Protein Folding In Vivo and In Vitro.* Washington, DC: American Chemical Society, 1993, pp. 72-83.

William C. Wimsatt, "Reductive explanation: a functional account", in: Robert S. Cohen (Ed.), *Proceedings of the Philosophy of Science Association, 1974.* Dordrecht, Holland: D. Reidel Publishing Company, 1974, pp. 671-710.

Alan C. Love
Department of Philosophy
Minnesota Center for Philosophy of Science
University of Minnesota
831 Heller Hall
271 19th Ave. S
Minneapolis, MN 55455
USA
aclove@umn.edu

Andreas Hüttemann
Philosophisches Seminar
Universität zu Köln
Albertus-Magnus-Platz
50923 Köln
Germany
ahuettem@uni-koeln.de