

DRAFT
NOT TO BE QUOTED

Andreas Hüttemann
Philosophisches Seminar
Universität Münster
Domplatz 23
48143 Münster
Germany
ahuettem@uni-muenster.de

and

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

Forms of Reductive Explanation in Biological Science:

Intrinsicity, Fundamentality, and Temporality[†]

[†] Order of authorship is alphabetical; both authors contributed equally to this paper. We are grateful to the comments and feedback we have received from many individuals (Ingo Brigandt, Tom Doyle, Susan Hawthorne, Marie Kaiser, Ken Schaffner, Ken Waters, Marcel Weber) and at various venues (Dept. of Philosophy, St. Cloud State University; Seminar on ‘Reductionism in Biology’ at the University of Minnesota; Dept. of Philosophy, University of Münster; Seminar on ‘Philosophy of Biology’ at the University of Münster; Center for Philosophy and Ethics of Science, University of Hannover; Minnesota Center for Philosophy of Science, Biology Interest Group at the University of Minnesota). We also 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.

Forms of Reductive Explanation in Biological Science:

Intrinsicality, Fundamentality, and Temporality

Abstract

The inapplicability of variations on theory reduction in the context of genetics and their irrelevance to ongoing research has led to an anti-reductionist consensus in philosophy of biology. One response to this situation is to develop accounts of reductive explanation (e.g. part-whole relations) that better correspond to actual scientific reasoning, thereby purchasing applicability and relevance. Working from this perspective we explore three different features, fundamentality, temporality, and intrinsicality, which arise from distinct aspects of reductive explanation: composition and causation. Focusing on these features generates new forms of reductive explanation and conditions for their success or failure relevant in biology and other sciences. This analysis is applied to the case of protein folding in molecular biology, which demonstrates its applicability and relevance, as well as illuminating reductive reasoning in a specific biological context.

1. Introduction
2. Composition, Causation, and Varieties of Reduction
 - 2.1 Composition versus causation
 - 2.2 Epistemology, ontology, and kinds of explanatory reduction
3. Part-whole Reduction: Intrinsicality, Fundamentality, and Temporality
 - 3.1 Intrinsicality and fundamentality
 - 3.2 Temporality
 - 3.2.1 Atemporal part-whole reduction
 - 3.2.2 Temporal (causal) part-whole reduction
4. The Protein-folding Problem
 - 4.1 Background and significance
 - 4.2 Reductive explanation in molecular biology
5. Philosophical Evaluation
 - 5.1 Application: Intrinsicality and fundamentality
 - 5.2 Relevance: Temporality
6. Conclusion

1. Introduction

The status of reductionism in the life sciences is a central issue in philosophy of biology . The most robust account of reductionism in the post-positivist context (i.e. second half of the 20th century) was Kenneth Schaffner's general reduction-replacement (GRR) model based on Ernest Nagel's discussion in *The Structure of Science* . Schaffner used this account to argue that reductionism was occurring in biology, especially as seen in the molecularization of genetics . But, importantly, this philosophical interpretation was not relevant to ongoing research methodology; it was not a primary aim of scientists to achieve this kind of reduction . The ensuing discussion about the relationship between classical and molecular genetics is expansive but largely characterized by a polarity between those in favor of reductionism and those against it . The so-called anti-reductionist consensus coalesced around both the inapplicability of the GRR model (or some modification thereof) and its lack of relevance to actual research, spurring the development of different ideas concerning the nature of theory structure . An oft-cited reason for this 'failure of reductionism', both in terms of application and relevance, is the misappropriation of philosophical models of reductive reasoning derived from the physical sciences .

One response to this situation was to develop different accounts of reductionism that were not subject to the criticisms of the GRR model and more explicitly sensitive to the reasoning processes present in genetics and molecular biology . Another similarly motivated response was to shift to different vocabularies such as 'mechanism' even though themes associated with reductionism remained (such as mechanism descriptions 'bottoming-out'). Both types of response share a criterion of adequacy that motivated the anti-reductionist consensus: any account of reductionism should be applicable and relevant to actual biology.¹ Instead of

¹ 'What geneticist could take seriously any explication of 'reductionism' which leads to the conclusion that

answering the question of reductionism in the negative, they sought to provide a perspective on reductive explanation that was minimally applicable if not also relevant to ongoing research. The philosophical task was to explicate reductionist reasoning in particular areas of biology and understand the criteria used by scientists to assess whether reductive explanation is successful, not be for or against reductionism *per se*. This criterion takes on increased significance when the diversity of reasoning practices in biological science is recognized. Reductive explanation exhibits many forms in different areas of life science, let alone in diverse areas of science such as physics or sociology. The philosophical task is thus much larger: to understand and explicate the diversity of reductionist reasoning practices in biology and other sciences ('application'), with special attention to their utilization in ongoing research ('relevance').

The aim of this paper is to contribute to this broad philosophical task of comprehending the heterogeneous nature of reductive explanation in biology. Our strategy is to distinguish two core concerns relevant to reductive explanation, composition and causation, and trace out the consequences for some different forms of reductive reasoning construed as part-whole relations. Three features of biological explanation emerge as especially salient: fundamentality, intrinsicality, and temporality. The latter two features have not received sufficient philosophical attention in analyses of reductive explanation and the reasons why are complex but superficially straightforward. First, part-whole relations considered spatially presume that parts are contained within the whole by some boundary, which encourages the perspective that parts are intrinsic (or internal) to the whole. Thus, intrinsicality is a precondition for discussing part-whole relations rather than an auxiliary element in their evaluation. Second, relations between parts and wholes are about spatial relationships that relate to composition and arrangement. Thus, temporal relations do not seem to be relevant. It is our contention that explicitly considering both of these

molecular genetics does *not* amount to successful reduction of classical genetics' .

factors along with fundamentality substantially augments our understanding of reductionism in biology.

We begin by highlighting the distinction between composition and causation. Reductive explanations often involve both claims about the *composition* of higher-level entities (wholes) by lower-level entities (parts) and the *causal* production of higher-level entities by lower-level entities. Next we review core themes in traditional theory reduction in order to argue that many ideas emerging from these debates are atemporal in nature (e.g. isolating correspondence relations or identities). As a consequence they do not adequately describe reductive explanations found in scientific reasoning that involve causal claims. Then we set out a basic framework for part-whole reductive explanations and begin to distinguish different explanatory forms using intrinsicity, fundamentality, and temporality. In keeping with our stated criterion of adequacy, we treat the applicability and relevance of our account in the context of an example of reductive reasoning in biology: the protein-folding problem. Our inclusion of temporality and intrinsicity in particular 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 are different. Additionally, it clarifies relations between these different conceptions of reduction and demonstrates that a failure of one form of reductive explanation does not imply a failure of reductionism *per se* or a failure of explanation.

2. Composition, Causation, and Varieties of Reduction

2.1. Composition versus causation

Within the sciences, there are at least two major features relevant to pursuing a reductive explanation. The first pertains to *composition*: higher-level entities are composed of, realized by,

or nothing but lower-level entities. An anatomical unit, such as 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. The second major concern involves *causation*: higher-level entities are caused by, 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 within a filament molecularly ratchet along another set of proteins. The appropriate folding of a particular myosin protein is determined by its amino acid sequence.

Many reductive explanations in the biological sciences are a mixture of compositional and causal claims (cf. Craver and Bechtel [2007]). There is no rigid observance of this distinction in reductions offered by biologists even though the two features can be isolated in retrospect. From a philosophical standpoint, keeping these features distinct is important for adequately reconstructing the reductive reasoning offered in biology (and other sciences). Consider the protein-folding problem in molecular biology. In *The Structure of Biological Science* ([1985]), Alex Rosenberg offered an interpretation of reductionism in the context of claims about protein folding.

The watchword of reductionism in biology has been the slogan that “function is a consequence of conformation, and conformation is specified by sequence.” That is, the “sequence” or linear one-dimensional order of the atomic components of a biologically significant molecule causally determines its three-dimensional structure, its shape or “conformation”; its three-dimensional structure causally determines its effects, and in particular all its biological functions (Rosenberg [1985], 73).

Thus far Rosenberg's claims about the reductive reasoning involved in protein folding research concern causation. This is indicated by key phrases such as “causally determines” and “is determined exclusively by”. Rosenberg continues:

The shape of the whole molecule [...] is determined exclusively by the order of the amino acids that compose it and by the chemical properties of these amino acids. That is, given a specification of the [...]

amino acids [...], we can deduce the shape of the whole molecule from the fact that some amino acids are hydrophobic – not water soluble, some are hydrophilic – water soluble, some are charged negatively and some positively, and that some contain larger and some smaller atoms in their side-chains (Rosenberg [1985], 75-6).

Here we see a shift to issues of composition, signaled by the language of ‘deducing’ the final protein fold. Given a particular set of amino acid residues, the composition of the protein, one can deduce the three-dimensional conformation of the folded protein. The *causal determination* of folding from the amino acid sequence in the polypeptide can be distinguished from deducing or inferring the native conformation from the amino acid sequence *composition* of the polypeptide. This is not to say that these two elements are explicitly kept apart in scientific discourse pertaining to explaining protein folding. But the question of whether native protein conformation can be inferred from the linear polypeptide composition is separate from the question of whether there is a *causal* explanation of how the folding occurs. An affirmative (or negative) answer to one does not imply an affirmative (or negative) answer to the other. The difference between these two is discernable in further comments offered by Rosenberg: “All we need ... is information about the chemical milieu and about the chemical bond already available in physical chemistry . Information about the chemical milieu does not concern the composition of the protein but seems relevant to a causal account of its folding, which involves interactions among the amino acid residues that compose it. These two subtly intertwined features need to be explicitly recognized in an analysis of reductive explanation.

Although numerous issues attend composition and causation in reductive explanation, we want to extract three key elements. The first is ‘intrinsicity’ and relates to composition. Claims about reducing a higher-level entity to its component parts involves a presumed individuation of the higher-level entity such that its components can be distinguished from other entities (a surrounding context). This is an epistemological sense of intrinsicity, rather than

one concerned with the metaphysics of intrinsic properties, because what counts as intrinsic depends on the explanatory goals of researchers. If the explanatory goal is to reduce a cell to its components parts then the boundary between intrinsic and extrinsic is often understood to be the cell membrane. If the explanatory goal is to reduce the heart to its component parts then the theoretically delimited boundaries of the organ establish the line between intrinsic and extrinsic. Epistemological intrinsicity is relative to the goals for a particular explanatory reduction.

The second feature is fundamentality and also usually relates to composition (cf. Sarkar [1998]). An intuitive interpretation of fundamentality corresponds to the assumption that higher levels of organization are reduced to lower levels of organization: parts are taken as more fundamental than the compound. This basic sense can be qualified so that the rationale for taking the parts as more fundamental is made explicit, such as a having a particular set of restricted properties (e.g. biochemical moieties). These kinds of qualifications are usually spelled out in the context of explanation where the relevant properties of the parts are specified. Qualified senses of fundamentality are important to distinguish because they can introduce considerations about intrinsicity over and above what is assumed in the intuitive sense. For example, a qualified fundamentality might pick out biochemical properties as explanatory of cell properties. But this allows for the possibility that biochemical properties of entities outside of the cell (i.e. extrinsic) are fundamental and the possibility that other intrinsic properties of cellular components (e.g. location) are not considered explanatory. Therefore, a reductive explanation might be characterized differently depending on whether it is assessed with respect to fundamentality, intrinsicity, or both simultaneously.

The third feature, which relates to causal aspects of reductive explanation, is temporality. Any explanation that is causal in nature will involve some element of temporal duration. If the

aim is to 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. From an epistemological vantage point, temporality pertains to the heterogeneous measures of time used by researchers. Different interval choices for temporality are utilized depending on the explanatory goals of the researchers and different reductive explanations of similar phenomena may have incompatible temporal measures, such as time measured in terms of event sequences or absolute chronology (SUPPRESSED FOR REVIEW).

Before we can observe the significance of incorporating intrinsicality, fundamentality, and temporality into a philosophical analysis of reductive explanation, it is necessary to see how previous discussions of reductive explanation have tended to overlook them.

2.2. Epistemology, ontology, and kinds of explanatory reduction

Although there are many meanings of ‘reductionism’ (SUPPRESSED FOR REVIEW), our focus here is epistemological reduction, and explanatory reasoning more particularly. Explanatory reduction often involves both compositional and causal claims, such as claims that certain kinds of entities (systems, properties, states, laws etc.) are to be identified with what we hitherto took to be different but—in some sense—more fundamental entities or certain phenomena (the behavior of systems) can be explained in terms of causes, laws, or properties (*inter alia*) that are part of a realm that is considered to be more fundamental. Explanations qualify as reductions when they satisfy certain constraints. With respect to compositional aspects of reduction, this involves ensuring constituents mentioned in the *explanans* bear appropriate spatial relations to the *explanandum* phenomena. In the case of causal relations, this involves ensuring that the causes,

laws, and/or properties mentioned in the *explanans* are appropriately (more) fundamental. Explanatory reduction is only one domain of epistemological reductionism.²

The shared background for discussions of reduction in philosophy of science is Nagel's account of the *formal* criteria of reduction. One of Nagel's primary concerns was whether an older theory (the reduced theory) is reducible to its successor (the reducing theory). Reduction was conceived of as a special case of deductive-nomological explanation. If the old theory reduces to the successor then the laws of the old theory as well as its observational consequences can be deduced from the successor theory. Successful 'Nagel reduction' integrates the old theory into the successor theory and provides a clear sense in which the successor theory is better than its predecessor. Reductionism in this sense is epistemological because it claims that one theory can explain another theory via deductive derivation. It is a reductive explanation because the explanatory theory is considered to be more fundamental.³ This approach provides a rationalization of the history of science but did not capture the lion's share of attention in subsequent literature (and we ignore this historical dimension as well).

What did exercise many philosophers was a question containing potential ontological consequences: the problem of bridge-laws. If in a putative Nagelian reduction the two theories invoke different terminology, as in the case of thermodynamics and statistical mechanics, connections need to be established that link terms such as 'temperature' or 'entropy' with notions from statistical mechanics. According to one influential strand of thought (e.g. Schaffner [1967], [1993]), these bridge laws ought to be construed as biconditionals that express synthetic

² Methodological reductionism is a label for various epistemological claims that it is a fruitful (or the most fruitful) heuristic strategy to investigate particular natural phenomena using reductionist methods. Explanatory and methodological reduction can be decoupled because a commitment to one does not entail the other; methodological reductionism does not guarantee explanatory success and a successful explanatory reduction does not imply that methodological reduction is the most favorable general strategy of inquiry.

³ The succession of theories in the history of science (e.g. a less general earlier theory is reduced to a more general later theory) has been labeled 'diachronic' reduction by some philosophers.

identities, such as between the temperature of an ideal gas and its microstate. The question of reduction translates into a question of whether there are such identities. This is an ontological claim that the properties described by the reduced theory are identical with those picked out by the reducing theory. Seeking identities along these lines can be a method of exploring relations between two theories both of which are currently accepted to some degree (cf. Wimsatt [1976]), such as folk-psychology and neuro-physiological accounts of the brain. If these identities can be secured then the Nagelian account provided a rationale for any resulting theory replacement.

Many concepts that proved to be central in reductionism debates were introduced in this context. For example, the fact that certain properties picked out by the theory to be reduced can be multiply realized by properties described in the reducing theory was taken to preclude the identification of higher-level and lower-level (or more fundamental) properties or types (cf. SUPPRESSED FOR REVIEW). The notion of supervenience seemed to provide some consolation and was meant to capture the idea that a particular instantiation of one set of properties of a system (e.g. physical ones) ‘in some sense’ determines another instantiation from a different set of properties (e.g. mental ones), even if a demonstration of identities among types remains beyond reach. Jaegwon Kim defines strong supervenience as follows: ‘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’ .

Three features of these concepts bear mentioning. First, synthetic identity, (multiple) realization, and supervenience build off of the *ontological* possibilities inherent in the problem of bridge laws within the epistemological approach to reduction originally offered by Nagel. Second, identity, realization, and supervenience focus on relations that obtain between two

properties (different in kind; e.g. mental and physical) of one and the same object or system.⁴ The question at issue is whether a putatively non-physical state or property of a system is identical to, or at least metaphysically dependent on, physical states or properties of the *same* system. Third, these concepts are fundamentally *atemporal*. Identity and realization are atemporal relations because they obtain strictly at a time slice.⁵ (A *temporal* relation is one in which a property or state at t is related to another property or state at t^* , with $t^* \neq t$.) This atemporality is consistent with the identity relation persisting over time, although this is usually assumed rather than articulated. In the above definition of supervenience it is the mental properties *at t* and the physical properties *at t* that are taken into consideration. It is either supposed that a certain property realizes another property only at a particular time or assumed that these identities between properties persist.

Much of the discussion about reductionism in biology over the past few decades has revolved around the relations between classical genetics and molecular genetics. The primary context in which these debates occurred is the problem of bridge laws and the (ontological) question of synthetic identities between different theories. This includes the presumption of intrinsicity, i.e. relations between different properties of one system, and atemporality, i.e. relations that obtain strictly at a time slice, such as identity, realization, and supervenience.⁶ Much of the so-called antireductionist consensus was forged in this same context, largely accepting this framing of the question. We contend that our understanding of reductive explanation in biology is amplified when these two assumptions no longer operate as presuppositions but are brought to the foreground for explicit consideration.

Although the question of whether predicates of different theories pick out identical

⁴ There is at least one exception for supervenience that proves the rule (see Hoffman and Newen [2007]).

⁵ Wilson and Craver ([2007]) broach the issue of whether realization is necessarily intrinsic and synchronic.

⁶ A few philosophers have noted the difference between atemporal and temporal types of reduction.

properties could be relevant and applicable to reductive modes of explanation in scientific reasoning, it appears to be of limited significance in biology and has not substantially clarified the putatively reductive relations between classical and molecular genetics. In parallel with discussions emerging out of Nagel's seminal work, a variety of authors have conceptualized reductionism as a question about the relationship between parts and wholes.⁷ This approach plays important methodological and explanatory roles in biology and largely bypasses the issues that arise in the context of theory reduction discussed above. For example, Sahotra Sarkar ([1998]) argues that reductions in biology (and elsewhere) are explanations of the phenomena of one non-fundamental realm (e.g. a whole) in terms of another more fundamental realm (e.g. the parts). These reductive explanations vary in their strength based on how they fulfill criteria pertaining to the representation of objects in a reductive explanation.⁸ Sarkar scrutinizes two forms of explanatory reductionism ('genetic' and 'physical') derived from this strategy in the context of genetics. For example, physical reductionism is the explanation of biological phenomena using the physical properties of constituent molecules and macromolecules; 'the behavior of wholes is supposed to be explained by those of their constituent parts'. We think this focus on part-whole reduction is broadly correct. The question we tackle in the following section is how the inclusion of temporality, fundamentality, and intrinsicity expands the forms of reductive explanation available for characterizing reductive reasoning within biology.

3. Part-whole Reduction: Intrinsicity, Fundamentality, and Temporality

Recall that the features of intrinsicity and fundamentality on the one hand and temporality on

⁷ Issues related to parts and wholes were present in Nagel's original discussion but overshadowed in subsequent developments of reductionism.

⁸ 'Representation' refers to how natural phenomena are symbolized, embodied, pictured, or designated such as in equations, scale miniatures, or abstract diagrams.

the other correspond to the distinct issues of composition and causation in reductive explanation, respectively. As noted, this can be initially explicated for part-whole explanatory reductions as follows: (i) for compositional relations between parts and wholes, intrinsicity refers to the fact that parts are presumed to be *contained within* or *internal to* the whole by some boundary or line of demarcation; (ii) fundamentality refers to either an intuitive sense of fundamentality (parts *qua parts* are more fundamental than the compound) or a qualified sense in which the parts are understood as more fundamental by virtue of having a certain restricted set of properties; (iii) for causal relations between parts and wholes, temporality refers to the fact that a property of a whole at t^* is explained in terms of properties of the 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^* (where $t^* \neq t$), then a causal relation is one in which a property or state at t determines or influences another property or state at t^* , where $t^* > t$.⁹ In the context of causal part-whole reductions, this is understood as the state or properties of the parts and their interactions at t determining or changing the state or properties of the compound at a later time t^* . Different forms of reductive explanation emerge when these three features are explicitly considered. We first focus on intrinsicity and fundamentality before turning to temporality in order to distinguish distinct perspectives on causal part-whole reductions in biology.

3.1 Intrinsicity and fundamentality

Part-whole reductive explanations considered with respect to composition are reductions in two respects. First, they appeal exclusively to *intrinsic* features of the compound system in question. No extrinsic features are invoked to explain the property of the whole. Second, they appeal exclusively to a more *fundamental* realm or lower-level features (the parts) as compared to the

⁹ We assume that properties of wholes at t^* cannot be explained in terms of the properties of parts at a later time.

compound considered as a whole (non-fundamental realm). Consequentially, attempted part-whole reductions in terms of composition can fail either because intrinsicity is violated, fundamentality is violated, or both (assuming a qualified sense of fundamentality).

Intrinsic properties have been variously characterized in metaphysics (see, e.g., Langton and Lewis [1998]), although the raw intuition is that ‘a property is intrinsic if anything that has it has it regardless of what is going on outside itself’ . If we investigate the epistemological issue of how scientists individuate systems from their environments (thereby deciding what is ‘inside’ and ‘outside’), then a specific characterization of what intrinsic and extrinsic means can be determined (though is unlikely to generalize). It is necessary in scientific reasoning for researchers to demarcate the system of interest from its environment—natural phenomena do not come with labels attached to indicate their boundaries. If this reasoning takes the form of a reductive explanation, then the determination of boundaries predisposes researchers to focus on particular causal factors. Once the system-environment distinction has been drawn and the system is decomposed into parts, explanations will tend to investigate the causal powers of the parts within the system (‘intrasystemic’ - intrinsic) rather than the role of aspects external to the system (‘intersystemic’ - extrinsic) . This observation has been deployed to emphasize how explanations of biological phenomena, such as ontogeny, may be biased toward attributing causal responsibility to intrasystemic features (e.g. genes) rather than intersystemic ones (e.g. predation) . What has gone unnoticed is that this bias is neutral with respect to whether extrinsic features also meet a qualified sense of fundamentality. For example, if a system is individuated from an environment and partitioned into parts that are fundamental in the qualified sense of possessing a particular set of properties, then it is not only the causal responsibility of systems or compounds at the same ‘level’ within the environment that are systematically ignored. Any role

played by the parts that these extrinsic systems are partitioned into will also be neglected, even though they also may be fundamental in the qualified sense. From the perspective of a part-whole reductive explanation this means the question of fundamentality is decoupled from that of intrinsicity because the failure of the latter does not imply failure of the former.

If the part-whole relation is considered atemporally, then a violation of intrinsicity seems impossible. The very idea of part-whole relations presumes that parts are contained within the whole by some boundary (i.e. nesting of parts). But if the part-whole relation is considered temporally, which means introducing causal relations in addition to compositional ones, then it is possible for intrinsicity to be violated because a part of a whole at time t may no longer be a part of the whole at t^* or *vice versa*. Consider a system S individuated from environment E and partitioned into components (s_1, \dots, s_n) at time t . Let S be the heart, E be the rest of the body cavity, and s_1, \dots, s_n be cells within the heart. If the cells happen to be blood cells coursing through nascent atrial chambers during ontogeny, then their presence in the heart at time t meets the intrinsicity condition even though at t^* they have passed out of the heart. This is significant because blood flow is a key factor in the proper shaping of the heart during embryogenesis (Hove et al. [2003]). Thus the blood cells coursing through the heart at t have a causal effect on the shape of the heart at t^* when these cells are no longer intrinsic to S . Notice that this is still a form 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). Alternatively, we may be interested in changes of heart rhythm, which are modulated by adrenergic hormones secreted in E . These hormones are extrinsic to the heart (S), even though they meet the fundamentality condition, and exert an effect through a temporally extended process. Another possibility is that other organs directly interact with the

heart to bring about changes in its morphology during embryogenesis (e.g. through physical contact). These other organs are not only extrinsic to *S* but also non-fundamental because they are at the ‘same’ non-fundamental level as *S*.

We can summarize these different combinations of meeting the three conditions in a table of reductive explanatory forms (Table 1). This displays not only a diversity of kinds of reductive explanation but also the heterogeneous conditions for how they can succeed or fail. The first two cases (I and II) correspond to those that can be recovered from previous discussions of reduction, with the latter including compositional relations extended through time. But the other cases are different. The third (III) corresponds to an extrinsic entity from a fundamental level explaining a system property (e.g. the adrenergic hormone case). Although it is reductive in the sense of appealing to properties at the fundamental level, it fails in the sense of a whole being causally explained by its parts. The fourth case (IV) is an example of one organ interacting with another organ (as described above), where it is not only that the entity is extrinsic to *S* but also that it resides at the same level of fundamentality. Here there is a failure of the whole to explain its parts and a failure of the fundamental level properties to explain non-fundamental level properties.

Table 1

<i>Forms of Explanation</i>	<i>Fundamentality</i>	<i>Intrinsicality</i>	<i>Temporality</i>
I	Y	Y	N
II	Y	Y	Y
III	Y	N	Y
IV	N	N	Y

Once temporality is admitted as a possible variable to consider, intrinsicality and fundamentality become independently interesting to analyze.¹⁰

¹⁰ The table omits the possibilities previously ruled out, such as atemporal part-whole reductions not violating intrinsicality or fundamentality, or fundamentality failing when intrinsicality holds. This latter situation captures the sense that if a feature is intrinsic to a system *S* then in order to be contained within *S* it must be instantiated at a more fundamental level than *S* itself.

It should again be stressed that what counts as intrinsic and extrinsic is relative to the explanatory aims involved in particular explanations. This sheds light on the ‘context objection’ to reduction, which claims that a reduction can be blocked because of an ineliminable appeal to contextual factors. The standard reductionist rejoinder is to pursue a reduction of this context, which is to stress maintaining the fundamentality condition even when intrinsicity is violated. But by distinguishing the two different forms of explanation where reduction can fail (III – intrinsicity fails; IV – fundamentality and intrinsicity fail) we are able to recover a more nuanced reading of the context objection. Even if the standard rejoinder is able to preserve fundamentality, the context objection retains validity *because* of the intrinsicity precondition for part-whole reductive explanation. If an attempt is made to recover intrinsicity by redrawing system boundaries (e.g. treat the entire body cavity as the system and then the adrenergic hormone is formally a part of this new whole), then we have changed the *explanandum*. To reconfigure the boundaries of a system changes what counts as the whole and its parts, regardless of whether you claim that they are ‘reducible’ or not. An evaluation of whether a part-whole reductive explanation succeeds or fails depends on how ‘wholes’ are individuated, which means any redrawing of the boundaries, even if epistemologically warranted by the demands of empirical investigation, constitutes a change of the original question about whether a whole can be reductively explained by its parts. Therefore we adopt as a default principal the system individuations utilized by the scientists.

3.2 Temporality

3.2.1. Atemporal part-whole reduction

Thus far we introduced temporality only to see its critical interaction with intrinsicity. Now we

explicitly consider temporality in the context of part-whole reductions. One way to seize on the significance of temporality in reductive explanation within biology, due in part to an interest in causation over and above composition, is to contrast it with what we find in physical science. Part-whole explanations in physics are primarily atemporal and what follows is a brief sketch of their characteristics .

A synchronic part-whole reduction is the explanation of the state of a compound system at a time t on the basis of the states the parts are in at the same time. For example, we might explain why a compound system (e.g. an ideal gas) has the determinate energy value E (the macro-state) by appeal to the determinate energy values of its constituents (e_1 to e_n ; the states of the parts). One can reconstruct this as a deductive-nomological explanation that relies on the states of the parts (particular facts) and a law of composition. The law of composition tells us how the states of the parts 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 in that it only allows states of the parts and a summation rule or law of composition in the *explanans* but it is also clearly atemporal. Note that if different sets of states of the parts give rise to the same state for the compound there is no obstacle for reduction. The explanation is not undermined by multiple realization because identity is not the pertinent issue.

Synchronic part-whole explanation pertains to the state of a compound system. Besides the state of a physical system one can also incorporate the dynamics or temporal evolution of the system to provide a complete description of its behavior.¹¹ As in the case of the state of the compound, the dynamics of a compound system is often broken down to that of the parts. For instance, the Hamilton operator for a compound system can be analyzed in terms of the Hamilton

¹¹ The dynamics of a classical physical system is characterized by Newton's second laws with a specific force function. In the case of quantum mechanics the dynamics is characterized through the Schrödinger-equation with a specific Hamiltonian.

operators for the parts along with interaction terms and a law of composition. The relation between the Hamilton operator for the compound and the Hamilton operators for the parts (i.e. a part-whole relation) is reductive and *atemporal* – as in the case of the states of the compound and the parts. It is *atemporal* because it is not the case that the Hamilton operator of the compound at time t^* is calculated in terms of the Hamilton operators of the parts and their interactions at an *another* time t . (Typically, the Hamilton operators of the parts and the compound are time-independent.)¹²

Although we have explicated the notion of *atemporal* part-whole explanation in terms of laws, as is appropriate for (fundamental) physics, it is possible to connect this vocabulary to biological discourse by assuming that laws are grounded in dispositions or causal powers. This allows the following formulation: for a system that is *atemporally*, part-whole reducible, its causal powers at t can be explained in terms of (reduced to) the causal powers of its parts (invoking rules that tell us how causal powers add up). Language about laws is thus translated into the language of causal powers. Alternative translations could be accomplished with different accounts of causal generalizations in scientific explanation. The forms of part-whole reductions characterized here do not rely on a specific, substantive commitment to the nature of explanation. (Our terminology reflects only an expository preference, which could be suitably translated in terms of these alternatives as needed.)

The export of *atemporal* models of reductive explanation from philosophical reflection on the physical sciences has hindered some discussions of reductive part-whole explanations in

¹² The reductive explanation of the dynamics of a compound in terms of the dynamics of its parts is usually associated with the explanation of the state of a compound in terms of an earlier state of the compound. To give an explanation of how the state of a compound evolves from t to t^* one needs the dynamics that describes this evolution (e.g. the Schrödinger-equation with a Hamiltonian for the compound). Reduction enters the picture because the dynamics of the compound is usually determined in terms of that of the parts. This kind of combined explanatory reasoning (later state in terms of earlier state; the dynamics involved in terms of that of the parts) has been called (maybe somewhat confusingly) “diachronic microexplanation”.

biology. Relations between parts and wholes in biological hierarchies have long been recognized as involving temporality and causal interaction or interdependence . The part-whole reductive explanation described above for a physical system essentially converts causal claims into compositional ones. A causal claim that a property of a whole at t^* is explained in terms of properties of the parts at an earlier time t is broken down into a question of the compositional relations between the whole and its parts at t^* and a question about the dynamics between the parts (from t and t^*). This implies that it is necessary to articulate compositional relations in order to advance reductive causal explanations. It is not clear this is appropriate for the hybrid compositional and causal claims found in biological explanations, because it is often the case that explanations of a system property at a later time (t^*) are sought in terms of differences in properties of the parts at an earlier point in time (t), without attempting to provide the compositional relations between the whole and its parts at the later time. This encourages treating temporal (causal) part-whole reductions separately.

3.2.2 Temporal (causal) part-whole reduction

Part-whole reductions (and explanations more generally) in biological science are often *temporal*. They are temporal in the following sense: a property of a whole at t^* is explained in terms of properties of the parts at an earlier time t . The properties or behavior of the parts at t *cause* the compound to have a certain behavior/property/state at a later time t^* . Temporal or causal part-whole reductions can be described differently depending on whether laws are available for the description of the behavior of the parts. In those parts of biology where physical laws play a role, causal part-whole reduction can be spelled out as follows: the behavior of a whole at t^* is explained in terms of the laws and initial conditions that pertain to certain

parts of the whole at t . If laws are not available or difficult to pin down, we can switch to talk about causal powers or dispositions: the behavior of a whole at t^* is explained in terms of the causal powers or dispositions of certain parts of the whole at t . Causal part-whole explanations are reductive in that they place certain constraints on the *explanans*, allowing only laws or causal powers of the parts (and maybe also compositional rules). Put differently, they require intrinsicality and fundamentality.

Examples of causal part whole reductions frequently appear in the biological literature, such as explaining muscle tissue activity (behavior of a system) by appealing to the properties of muscle cells composing the tissue (parts of the system), which contain special motor proteins (e.g. myosin) that mechanically achieve contraction by molecular ratcheting . Temporal part-whole reduction explains the state of a compound or whole (muscle contraction) in terms of states/laws/dispositions of the parts at earlier times (myosin ratcheting) and must be distinguished from atemporal part-whole reduction, which explains the behavior of a compound in terms of the states/laws/dispositions of the parts at the same time. Atemporal part-whole reduction focuses primarily on *composition* whereas temporal part-whole reduction focuses on *causation*. How these different kinds of reductive explanation relate to one another is an interesting question in its own right.¹³ But it is enough here to highlight that, from an epistemological vantage point, atemporal part-whole reduction corresponds more appropriately to physical science reductive explanation and temporal (causal) part-whole reduction more accurately captures biological science reductive explanation. And, even more importantly, part-whole reduction must be distinguished from identity-reduction, which focuses on the relation of two kinds of properties/states that pertain to the same system rather than explaining properties of

¹³ We leave open the possibility that biological temporal part-whole reductions might eventually be explicated in terms of atemporal part-whole reductions in physics (according to the standards of the scientific community).

a compound system in terms of the properties of different systems (i.e. its parts). These differences make clear why multiple realization and supervenience are not easily transferable from the context of identity-reduction to temporal part-whole reduction. Even attempts to define supervenience for the part-whole relation are not applicable to causal part-whole reductions because they concern fundamentally atemporal relations. Causes do not *realize* their effects and the existence of multiple pathways from the causal powers of parts to a later state of the compound does not undermine causal part-whole reduction.¹⁴

It may be objected that part-whole explanations are not different from explanations of one property of a system in terms of another property of the same system. Consider an atemporal case: a property of the whole, the temperature of an ideal gas, is explained in terms of properties of its parts, suitably combined or summed (mean kinetic energy). What explains the property of the whole, strictly speaking, is the property (of the whole) to have such-and-such a configuration of parts. How does this differ from identity reduction where a property of the whole is explained in terms of another property of the whole? In dealing with this objection it is important to distinguish two issues. With respect to the gas we can ask two different questions: (i) Is it possible to explain why an ideal gas with a certain mean kinetic energy has a certain temperature? (ii) Is it possible to explain why an ideal gas has a certain mean kinetic energy on the basis of the behavior of the molecules of the gas? Question (i) concerns two kinds of behavior/properties of the same system, namely of the ideal gas as a whole. It is the question of why a system that has a certain kinetic energy at the same time also has a certain temperature. In

¹⁴ A multiplicity of causal pathways from parts to compounds would, however, undermine the attempt to reduce special science causal *laws* to physical science causal *laws*. Jerry Fodor discussed whether (1) $S_1(x) \rightarrow S_2(x)$ ('all S_1 situations bring about S_2 situations'), a special science causal law can be reduced to (2) $P_1(x) \rightarrow P_2(x)$, a physical science causal law. The reduction of (1) to (2) is in our terminology an identity issue. If the cause $S_1(x)$ is multiply realized (different constituents p_1 to p_n and p'_1 to p'_m), then this would undermine the reduction of (1) to (2), in the sense discussed in section 3.1. It would not undermine temporal part-whole reduction of $S_2(x)$ at t^* in terms of the causal powers of either p_1 to p_n or p'_1 to p'_m at t (with $t^* > t$). In short, multiple realization is an interesting empirical phenomenon whose significance for reductive explanation depends on the meaning of 'reduction' in use.

other words, question (i) asks for an explanation in terms of a bridge law. Question (ii), on the other hand, concerns the relation between the behavior of the components or parts of the ideal gas and the behavior of the ideal gas as a whole; i.e. the question how the individual kinetic energies add up to the kinetic energy of the whole. Thus, mean kinetic energy is a property of the gas as a whole and the explanation of temperature for the ideal gas in terms of the mean kinetic energy of the gas is a question concerning two properties of the same system. But there is another issue, namely the relation of the mean kinetic energy of the gas as a whole to the kinetic energies of the molecules, which pertains to the relation of parts and wholes, not the relation of properties within one and the same system.

In addition to the differences already adduced between atemporal and temporal part-whole reductive explanation, we should be wary of assuming that temporal part-whole reductions can be subsumed under a completely atemporal account of part-whole reduction. This would collapse cases II, III, and IV (from Table 1 above), which appear phenomenologically in scientific reasoning, into the situation of case I. Although there might be metaphysical motivations to do so, the epistemological criterion of applicability and relevance to scientific research encourages preserving these distinctions in order to better limn the contours of reductive reasoning in biology and other sciences. The default stance should be to recover the actual reasoning patterns in the science rather than offer a reconstruction that intentionally deviates from stated choices of spatial and temporal individuation. We have already observed that bringing intrinsicity (composition) and temporality (causation) into the foreground produces new insights on the traditional ‘context’ objection to reduction in terms of the different ways reduction might fail and the differences between reductive explanation in biology and physics. Now it is time to see their applicability and relevance to a specific example from biology.

4. The Protein Folding Problem

4.1 *Background and significance*

Researchers working on various aspects of protein folding have explicitly recognized the centrality of reduction in the context of their inquiry.

One reason for the fascination of the protein folding problem is that it represents an unusually concrete and limited case of the whole problem of reductionism. An unfolded protein is clearly a chemical object: a backbone of exactly repeated simple units each with one of an alphabet of 20 possible side chains. Its properties are relatively dull and quite predictable by summing up the properties of its components. A folded protein on the other hand, in addition to complexity and unpredictability, has acquired meaning: unity, controlled interaction with other systems, and biologically significant function. In many cases one can watch a protein undergo the spontaneous transition from randomness to directed functionality in the pace of a few minutes. There is apparently no extra information hidden within the starting state, so we feel that understanding the rules of the transformation would teach us worthwhile lessons about hierarchical organization, cooperative properties, and exactly how an organic whole becomes so much more than a sum of its parts .

The philosophical issue of reductive explanation is so tangible that it appears in textbook presentations of protein folding: ‘the function of a protein [...] is an emergent property resulting from exquisite molecular order. [...] We have taken a reductionist approach in dissecting proteins to their four levels of structural organization’ . It is also relevant in the present philosophical context because Sarkar poses the protein folding problem as an outstanding question in molecular biology pertinent to assessments of part-whole explanatory reduction .

The standard biological claim that correct folding is solely a function of the linear order of amino acid components in a protein might fail in two ways: (1) given the laws of macromolecular physics, the linear order of the amino acid residues may be insufficient to explain the three dimensional conformation of the folded protein, and/or, (2) the laws of macromolecular physics themselves may be insufficient to explain the three dimensional conformation of the folded protein (pp. 169-170). Only (2) would count as an unqualified failure of ‘physical reductionism’

for Sarkar, whereas if (1) obtains then ‘other’ macromolecular physical processes may provide the necessary explanatory supplement. Our inclusion of temporality and intrinsicity alongside of fundamentality provides further philosophical resources for interpreting this particular example of reductionist reasoning.

4.2 Reductive explanation in molecular biology

Proteins are composed of amino acid components (‘residues’) chemically linked by covalent peptide bonds together into a chain (‘polypeptide’). This linear chain is produced from a process termed translation whereby particular cellular constituents (ribosomes, themselves proteins) ‘translate’ a linear stretch of RNA with a triplet code of nucleotides (e.g. AAG), which indicate particular amino acid residues (AAG = lysine), into a linear polypeptide (‘primary structure’). Secondary structure refers to repeating patterns of coiling or folding (e.g. α helices or β pleated sheets) that obtain for a polypeptide as a consequence of regular hydrogen bonding of its generic structure; i.e. not due to the specific chemical moieties of each amino acid residue. Nearly all proteins adopt a three-dimensional structure (‘tertiary structure’) in order to be functional. This conformation of the polypeptide is understood in terms of specific interactions among its particular amino acid residues (e.g. hydrophobic interactions where certain residues avoid interaction with surrounding water by segregating to internal regions). The protein-folding problem (hereafter, PFP) concerns explaining how this active conformation is achieved for polypeptides subsequent to translation from RNA in the cellular context.¹⁵ ‘A functional protein is not *just* a polypeptide chain, but one or more polypeptides precisely twisted, folded, and coiled into a molecule of unique shape’. The linear sequence hypothesis (hereafter, LSH) holds that

¹⁵ The PFP is not isomorphic to the question of protein function because the latter is often due to ‘quaternary structure’, which refers to the aggregation of more than one tertiary structured protein.

the three-dimensional *folding* of a protein occurs as a consequence of only the properties of the amino acid residues in the polypeptide and their interactions. The folded protein is explained by the chemical interactions of its component residues as ordered in a linear polypeptide—the whole is a ‘sum’ of the interaction of its parts.

The LSH, as a putative solution to the PFP, claims that the primary structure of the polypeptide contains all of the information required for achieving the tertiary structure corresponding to three-dimensional conformation. There is an ambiguity in this statement of the LSH. One interpretation focuses on the phrase ‘contains all of the information required’ and sees the LSH as a claim about being able to *infer* the three dimensional structure of a protein from its linear sequence of amino acids. A second interpretation concentrates on the phrase ‘achieving the tertiary structure’ and takes the claim to be about understanding the kinetic, thermodynamic, and structural aspects of the protein folding process. The first interpretation construes the PFP in terms of prediction from composition, *that* you can make the correct inference (‘predictive construal’); the second interpretation glosses the PFP as a causal question of *how* the folding actually occurs (‘folding construal’).¹⁶ It is the latter sense that Sarkar intends in posing the question about the adequacy of physical reductionism and the one we focus on here. Whether a protein mechanistically folds only as a consequence of its amino acid residues is a causal question about reductive explanations of wholes in terms of parts.

The (ambiguous) LSH began to crystallize in the wake of the unraveling of the genetic code. Biochemists held that there were yet-to-be discovered rules that governed three-dimensional protein conformation, such as regularity of amino acid sequence. This expectation was a form of reductionist reasoning (cf. Jaenicke 2005]) and its first expression came from

¹⁶ One way these two interpretations are conflated is to talk of the ‘translation of information’ from the linear amino acid sequence (parts) into the three-dimensional conformation of the protein (whole).

Francis Crick: ‘It is of course possible that there is a special mechanism for folding up the chain, but the more likely hypothesis is that the *folding is simply a function of the order of the amino acids*’. Crick explicitly perceived himself as pursuing reductive explanations (Crick [1966], p. 10) and philosophical commentators saw protein folding as susceptible to the putative trend of reductionism in biology via molecularization .

Some of the strongest evidence in favor of the LSH was derived from experiments in the laboratory of Christian Anfinsen on the denaturation and refolding of ribonuclease proteins *in vitro* . Ribonucleases subjected to denaturing conditions were able to refold rapidly into the proper configuration rather than the myriad of other biochemical possibilities.¹⁷ The LSH was seemingly confirmed because this protein refolded into a favored thermodynamic conformation in the experiment purely as a function of the linear sequence of amino acid residues composing the polypeptide.

The three-dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metals ions or prosthetic groups, temperature, and other) is the one in which the Gibbs free energy of the whole system is lowest; that is, that the native conformation is determined by the totality of interatomic interactions, and hence by the amino acid sequence, in a given environment .

But the precise role of extrinsic factors in the folding process (‘normal physiological milieu’ and ‘in a given environment’) was not systematically explored even though Anfinsen’s group found that 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).

In recent research the causal role of extrinsic factors in protein folding has increasingly come to the fore. Although it has been claimed that ‘today’s view [is] that *whatever* its amino

¹⁷ ‘A chain of 149 amino acid residues [...] would be able to assume on the order of 4^{149} to 9^{149} different conformations in solution. The extreme rapidity of the refolding makes it essential that the process take place along a limited number of “pathways” . This is known as the ‘Levinthal paradox’ : how is the native protein conformation rapidly ‘determined by the totality of interatomic interactions’ if a random search through the vast array of possible biochemical pathways is impossible? .

acid sequence, a polypeptide chain will *spontaneously* fold into a stable configuration', the situation is more complex. First, many denatured proteins do not refold as cleanly as those originally studied by Anfinsen's group.¹⁸ Second, 'spontaneity' of folding is critically dependent on the 'normal physiological milieu', which includes more than just physico-chemical components of the environment.¹⁹ Chaperone proteins guide folding during and after polypeptide synthesis, as well as in refolding subsequent to stressful conditions such as heat shock.²⁰ Third, the term 'spontaneous', which appears routinely in descriptions of protein folding, should not be confused with instantaneous.²¹ There is a (rapid) temporal dimension in protein folding and it is not a 'static' achievement. The conformation necessary for functionality is stable but dynamic, fluctuating around one or more preferred states, and is affected by molecular interactions such as ligand binding or phosphorylation.

One of the reasons that molecular chaperones must provide oversight in the process of protein folding is that the cellular environment is highly crowded. Research on molecular chaperones has uncovered a complex set of causal roles for biological macromolecules in the

¹⁸ 'Many cell biologists, having been taught that polypeptide chains can spontaneously fold to the native state, have been frustrated to discover that, although spontaneous folding can occur for small simple proteins [...], spontaneous, high-yield folding to the native state might be the exception, rather than the rule'.

¹⁹ Relevant extrinsic factors are sometimes interpreted as solely physico-chemical. 'Protein conformation also depends on the physical and chemical conditions of the protein's environment. If the pH, salt concentration, temperature, or other aspects of its environment are altered, the protein may unravel and lose its native conformation'.

²⁰ 'Proteins need the assistance of molecular chaperones and folding enzymes to reach their native structure efficiently without formation of (large) aggregates'.

²¹ E.g., 'A polypeptide chain of a given amino acid sequence can spontaneously arrange itself into a three-dimensional shape'. 'Folding and association of nascent or refolding polypeptide chains are spontaneous and autonomous processes'. The adjective 'spontaneous' usually denotes something happening on its own, apart from external stimulus, or to denote a process that occurs when conditions are 'just right'. In different terminology, it might be said that a protein has a disposition to fold under appropriate environmental triggering conditions. The difficulty is that chaperone proteins are in many cases specific to the manifestation of proper folding and operate over extended durations of time; i.e. chaperones don't appear to be 'triggers' and specific ones are required to have 'appropriate conditions'. This is very different from the many non-specific triggers one could use to manifest the fragility of a glass window.

physico-chemical milieu.²² Distinct functional sets of chaperone proteins monitor protein folding during *de novo* synthesis, quality control, and the response to stress in eukaryotes . Chaperones function in different ways, as well as cooperatively , sometimes providing a sequestered environment (as a ‘nanocage’) for folding to occur, and at other times actively facilitating folding (Ellis [1998]).²³ Altering the volume of the cavity inside a nanocage experimentally increases the folding speed for small proteins by modifying the ‘energy landscape’ of the protein . 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 (Tang et al. [2008]). Even when mutations are introduced that lead to altered amino acid components in a polypeptide, which should as a consequence prevent correct folding, correct folding can be induced by the overproduction of molecular chaperones .

5. Philosophical Evaluation

Although this only scratches the surface of research into protein folding, we now apply our earlier analysis to reconstruct the forms of reductive reasoning in the LSH and the fate of reductionist explanations of protein folding. Causal (temporal) part-whole reduction and intrinsicity allow us to revisit Sarkar’s claim that protein folding is a potential failure of part-whole reductionism. Sarkar described two ways in which the LSH might be false: (1) Given the laws of macromolecular physics, the linear order of the amino acid residues is insufficient to explain the three dimensional conformation of the folded protein; and, (2) The laws of

²² ‘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’ .

²³ ‘Unlike [nanocages], GimC/prefoldin plays a more active role in protein folding by interacting with unfolded proteins and stabilizing them against aggregation for subsequent folding’ . ‘CCT therefore acts as a folding nanomachine that uses the conformational changes undergone by the apical domains [...] to force the folding of the two cytoskeletal proteins [...] by pushing together the two domains that were previously separated and interacting with two opposed regions of the chaperonin cavity’ .

macromolecular physics may be insufficient to explain the three dimensional conformation of the folded protein. We intentionally ignore (2) because it is not the focus of the LSH and concentrate on (1) to observe causal (temporal) part-whole reductive explanations *in situ*.²⁴

Explanations become reductionist when they conform to certain constraints. What these constraints are may vary and therefore what is taken to be a failure of reductionism in one context may not count as a failure in another. Two constraints play a role in the context of the attempt to explain protein folding in terms of linear sequence. The first constraint is that only properties of the parts are relevant for part-whole reductive explanation of protein folding (intrinsicity). Contextual or extrinsic causal factors are not supposed to play a role. No other causal factor contributes to the folding besides the amino acids and their interactions. The second concerns what counts as the fundamental or reducing realm . In this case, the fundamental realm is composed of the causal powers of the amino acids (‘parts’), as well as macromolecular laws that describe their interactions. Systems that have causal powers due to a complex three-dimensional structure that are absent from the linear polypeptide do not count as fundamental. This is a qualified sense of fundamentality in that the properties of the fundamental realm do not include complex three-dimensional protein structure.

5.1 Application: Intrinsicity and fundamentality

The folding construal of the LSH is a causal (temporal) part-whole reductive explanation.

Tertiary structure of a three-dimensional protein whole is explained by the causal interaction of its component parts at earlier times; the amino acid residues causally interact to bring about the

²⁴ (2) might fail if the macromolecular laws of interaction turn out to be false. But attempts to confirm the LSH are not concerned with the sufficiency of the laws of macromolecular physics. The main concern is the explanatory adequacy of the linear amino acid structure with respect to the native conformation of the folded protein *assuming* the macromolecular physics that governs macromolecular interactions.

state or property ('being appropriately folded') in the tertiary structured whole protein. In the context of the PFP, a property is intrinsic to the linear polypeptide if it is a property of one of its amino acid components (e.g. hydrophobic side chain). (Relational properties between amino acid residue parts count as intrinsic from this point of view.) Extrinsic properties will be anything in the 'normal physiological milieu', inclusive of physico-chemical components, such as H₂O, as well as other proteins (e.g. chaperones) or nucleic acids (e.g. RNA). This reading fits standard accounts of the LSH: 'The specific function of a protein [or functional shape] is an emergent property that arises from the architecture of the molecule [...] the information for building specific shape is intrinsic in the protein's primary structure'. The LSH is a claim about how the causal powers of a linear polypeptide confer a disposition to fold into a three-dimensional protein (tertiary structure); the disposition manifestation of folding is purely a function of the intrinsic properties (causal powers) of the linear polypeptide.

At this point the significance of molecular chaperones becomes more perspicuous because if their molecular activity contributes necessary properties required for folding, as opposed to merely appropriate environmental background, then the folding construal of the LSH is false. We think this is a felicitous reconstruction of the research into the role of chaperone assistance in protein folding; the intrinsic properties of the linear polypeptide arising from its amino acid residue parts are not sufficient for the protein folding manifestation in the cell. Chaperones do not merely trigger the initiation of a disposition manifestation but specifically and substantially contribute to its manifestation. Use of the phrase 'spontaneous folding' is either shorthand or refers to the predictive construal of protein folding. The *causal process* of folding not only requires appropriate environmental conditions but also the contribution of extrinsic chaperones; i.e. there is a failure of intrinsicity in a causal part-whole reduction. Additionally,

the causal contribution of the molecular chaperones in the explanation of the protein folding is a consequence of three-dimensional structure; a kind of property that the amino acid parts in question clearly lack. This implies that the current best explanation of protein folding is also a failure with respect to the fundamentality constraint on causal part-whole reductive explanations (case IV in Table 1). Systems with causal powers due to complex three-dimensional structure, rather than systems that lack such a structure, are necessary to bring about the native conformations of proteins *in vivo*.

It is critical to recall that there are two ways for a temporal part-whole reduction to fail: either the behavior of an *extrinsic* element from the fundamental realm is necessary (intrinsicity fails) or the behavior of an extrinsic system with a *non-fundamental* property, e.g. a complex three-dimensional structure is necessary (intrinsicity and fundamentality fail).²⁵ The former would obtain if the relevant extrinsic properties were not folded proteins but rather free floating amino acid residues or unfolded polypeptide chains (elements of the fundamental realm). But the operation of molecular chaperones is indicative of the latter and stresses that another extrinsic system with a non-fundamental property (a folded, tertiary-structured protein) is required for the parts of a new linear polypeptide chain to become a proper protein. Causal powers of systems with non-fundamental properties rather than causal powers of the systems that have fundamental properties only are necessary to bring about 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.²⁶

²⁵ Recall that fundamentality alone cannot fail because if intrinsicity holds then the parts must be instantiated at a more fundamental level than the whole (see footnote 10).

²⁶ Notice that this claim is epistemological because it relies on the individuation and decomposition of the system offered by scientists. 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 a metaphysical sense rather than causal part-whole reductive explanation. Here we remain studiously agnostic on the matter.

It might be thought that the rebuttal to the context objection to reduction (‘just reduce the context also’) is relevant here. For example, one can always shift to a larger whole that contains the protein and the crowded cellular milieu, such as the cell, thereby including the molecular chaperones. But even if a causal part-whole explanation of the behavior of the *cell* is feasible in terms of the intrinsic causal powers of its parts, this would change the question of what parts and wholes are being causally reduced. If this larger whole can be reductively explained in the sense under discussion it would still remain a fact that the folding of the protein cannot be causally explained in terms of its amino acid parts. With respect to the ‘smaller’ whole (the protein) there is a violation of intrinsicity, as well as a violation of fundamentality. The failure with respect to fundamentality does not disappear in the original explanation if we have an explanation for the larger whole that meets these constraints. The extrinsic features or systems appealed to over and above the causal powers of the parts in the original explanation are themselves not located in the fundamental realm. Shifting to a larger whole would lead to a causal part-whole explanation of a different system and does not undermine our assessment of the original explanation.²⁷

Although a complete explanation of the process of protein folding seems to be on the horizon, this explanation does not fully meet the expectations of early researchers who put forward the LSH. Researchers now explicitly recognize that relevant causal powers are not located intrinsically, within the parts of the linear polypeptide. We think it unlikely that researchers will consciously flag this as a failure of reductionism, but the extrinsic, non-fundamental causes in the folding process demonstrate that the LSH has not been fully

²⁷ Some have argued that reduction can be preserved by ‘extending the mechanism’ backwards (or forwards) in time. For example, molecular chaperones are composed of ‘parts’ and therefore we can first ‘reduce’ the operation of a chaperone protein ‘whole’ to its parts. This strategy is inappropriate in two ways. First, in this case, the concern is with proximate causes most closely related in a temporal sequence to the generation of folded proteins (including the activity of chaperones), and not distal causes that allow us to ‘bottom out’ with parts rather than wholes. Second, and more important, chaperone proteins require other chaperone proteins for their own proper folding so extending the mechanism backwards in time leads to a type of explanatory regress.

vindicated. Causal contributions are required from extrinsic molecular chaperones, which themselves are tertiary-structured proteins that have undergone correct folding, in addition to the intrinsic properties of the amino acids and their interactions.

It might be objected that what we have described is not a grave kind of failure because all the factors that go into the explanation are non-mysterious (all entities under discussion here are macromolecules of different types). While understandable this objection misses the point that the original reductionist hypothesis for protein folding was one cast in terms of intrinsic properties of the amino acid residues ('parts'). The failure of the LSH concerns the inability of causally explaining the folding process in terms of intrinsic properties of the parts of the linear polypeptide. Because only macromolecules are involved, the 'unintuitive' failure of reduction in terms of temporal part-whole relations allows one to clearly see why a failure of reduction is not necessarily a failure of explanation, nor a capitulation to enigmatic non-physical factors.

5.2 Relevance: Temporality

The conceptual distinctions and tools we have developed would retain their usefulness even if the process construal of the LSH had been vindicated.²⁸ A further advantage of our account of reductive explanation is a perspicuous reconstruction of the scientific discourse regarding protein folding.²⁹ The inclusion of temporality facilitates an interpretation of seemingly paradoxical statements made by protein folding researchers, such as the following:

At first glance, the concept [of chaperone assisted folding] is in conflict with the paradigm of autonomous,

²⁸ These tools have significance beyond the issue of reductive explanation because of the importance of temporality in biological explanations *per se*, something also explicitly recognized by biologists working on protein function. "The dream is to 'watch' proteins in action in real time at atomic resolution. This requires addition of a fourth dimension, time, to structural biology so that the positions in space and time of all atoms in a protein can be described in detail" (Henzler-Wildman and Kern [2007], p. 964).

²⁹ Our analysis also directly connects with the vocabulary used by researchers: 'there is a need for molecular chaperones because the intrinsic properties of proteins assure that incorrect interactions are possible' .

spontaneous protein folding established by Anfinsen and many others. The apparent contradiction is that if the acquisition of the unique three-dimensional structure of a protein is governed by its amino acid sequence and the resulting interactions between amino acid side chains, there should be no need for molecular chaperones. However, additional factors have to be taken into account. [...] Molecular chaperones do not provide steric information for the folding process and thus do not violate the concept of autonomous folding. (The information for folding is encoded solely in the amino acid sequence.)³⁰

Once the ambiguity between the predictive and folding construals of the LSH is recognized (see above, Section 4.1), we can see why it could simultaneously seem like chaperones threatened to falsify it but that the amino acid sequence still ‘contains all the relevant information’. Chaperone assisted folding is in conflict with the LSH construed as a causal process *in vivo* because it is a claim about causal part-whole reduction. The key role played by extrinsic, non-fundamental entities means that the causal powers of the amino acid residue parts are inadequate to bring about appropriate folding in the cellular context.

But chaperone assisted folding is not in conflict with the predictive construal of the PFP, which is not a claim about causal part-whole reduction. Rather, it concerns the ability to infer or predict the native conformation of a protein from the linear polypeptide composition rather than causally explain the process of folding.³¹ Thus, our analysis of causal part-whole reductive explanation helpfully reconstructs the discourse of scientists working on the PFP and indicates why the failure of the folding construal of the LSH due to chaperone assisted folding is

³⁰ This tension has been present for more than a decade. ‘The early observations made by Anfinsen and colleagues that denatured ribonuclease refolds spontaneously in the absence of any other proteins into an active enzyme, formed the basis for the [...] hypothesis [...] that all the information required to assemble a polypeptide chain into a biological three-dimensional structure is contained within the amino acid sequence [...]. Results of studies in a number of different experimental systems have recently led to the realization that protein assembly *in vivo* is more complex than was originally thought, and requires the involvement of other proteins that have collectively been termed molecular chaperones’. Some researchers disagree with the claim that all chaperones provide no steric information (e.g. Ellis [1998]).

³¹ We are ignoring the thorny question of what is meant by ‘information’ in the predictive construal of the LSH (see Godfrey-Smith and Sterelny [2007] for discussion). One reason why the predictive construal of the LSH seems successful despite the failure of causal part-whole reduction is due to the fact that the former *assumes* the extrinsic features in the models used to predict three dimensional conformation. This is observable in Anfinsen’s statement of the thermodynamic hypothesis (see above, Section 4.2), as well as in more recent research. For example, some models that try to predict the ‘kinetics’ of protein folding account for extrinsic causal contributions (e.g. denaturant concentration) by adjusting pre-exponential factors (Fierz and Kiefhaber [2005], p. 840).

completely compatible with the success of increasing successes associated with the predictive construal. Robust inferences of three-dimensional conformation from linear sequence information are possible even in the absence of particular details about the actual causal (temporal) process of protein folding in the cell and those researchers focused on the predictive construal of the LSH do not deny that extrinsic molecular chaperones are causally relevant for protein folding *in vivo*.

Our account of different forms of reductive explanation also has the added advantage of providing increased comprehension of philosophical claims about reduction in biology. Recall Rosenberg's specific claims about protein folding (see above, Section 1). The ambiguity of the LSH is contained within Rosenberg's reductionist reconstruction, indicated by the switch between *causal* (temporal) *determination* of folding from the amino acid sequence in the polypeptide and *deducing* or inferring the native conformation from the amino acid sequence composition of the polypeptide. Causal (temporal) determination involves causal part-whole reduction, which has not been borne out in protein folding research. But deduction of protein conformation using amino acid sequence constitution information corresponds to the predictive construal of the PFP. The question of whether native protein conformation can be inferred from the linear polypeptide is distinct from the question of whether there is a *causal* part-whole reduction of how the folding occurs. An affirmative answer to one does not imply an affirmative answer to the other. The failure of causal part-whole reduction does not entail non-supervenience (or non-identity) for the folded protein—no controversial mereological claim must be made (e.g. the folded protein is something over and above its amino acid components).

6. Conclusion

The goal of this paper was to generate an account of part-whole reductive explanation that was both applicable to actual scientific reasoning and relevant to ongoing research. The applicability is illustrated through the different forms of reductive explanation that were isolated by explicitly considering intrinsicity, fundamentality, and temporality. Although these features were arrived at in part through attention to the development of philosophical thought about reduction over the past few decades, they are also explicitly motivated by the reasoning practices of working scientists. The relevance is observable in the way our analysis usefully interprets the folding construal of the LSH and keeps the predictive construal separate. This implies that explicating these different forms of reductive explanation is methodologically useful for categorizing and prosecuting further research. Our account's relevance is especially poignant since a large number of diseases arise from the misfolding and aggregation of proteins (Dobson [2003]), which is tackled by the folding construal of the PFP where causal part-whole reduction is appropriate. The forms of reductive explanation also capture differences in reductive reasoning between physical and biological science while explaining why the appropriation of 'atemporal' physical science reasoning for comprehending 'temporal' biological reasoning is problematic.

For too long reductionism debates in the philosophy of biology have focused on theory reduction and its formal strictures. Both the anti-reductionist consensus and attempts to argue for reductionism in terms of supervenience fall into this tradition. We have argued that a conception of reduction that focuses on part-whole relations, supplemented with considerations of temporality, fundamentality, and intrinsicity, is more fruitful in comprehending reduction within biology. Analyzing the PFP from the perspective of causal part-whole reduction precisely apprehends the sense in which the folding construal of the LSH is a failure of reductionism, even though it is not a failure of reductionism in the more traditional sense.

The arc of reductionist research in protein folding investigations conforms to general claims about the development of scientific research programs . An initial choice is made to divide a system, the linear polypeptide, from its environment and then research focuses on the entities and interactions that are *intrasystemic* (the amino acid residues) to explain protein folding. Simultaneously, environmental conditions are simplified (e.g. *in vitro* denaturation experiments). Although it would be worthwhile to catalog the precise correspondence of different reductionist research heuristic biases within protein folding research, we only want to draw attention to how researchers responded to their lack of success—they learned from it. Reductionism is honored in the breach; the failure of causal (temporal) part-whole explanation occurred by way of prosecuting methodological reductionism.³² The failure of one kind of reductive explanation in the PFP arises from the bias against perceiving *intersystemic* features deemed extrinsic in the original demarcation of the system (molecular chaperones) as causally critical to the process of folding. Thus, the failure of reductionism detailed here did not arise from *antireductionism*, metaphysical or otherwise, but occurred precisely because a relatively successful non-reductive explanation of how proteins fold *in vivo* is now within our grasp.

³² ‘We tend to start with simple models of complex systems—models according to which the parts are more homogeneous, have simpler interactions, and in which many differentiated parts and relationships are ignored. [...] But then as our models grow in realism, we should both capture more properties and see more of them as organization dependent’ .

