6-2017

Aspects of Biological Explanation

Derek J. Skillings
The Graduate Center, City University of New York

ASPECTS OF BIOLOGICAL EXPLANATION

by

DEREK SKILLINGS

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2017

Aspects of Biological Explanation

by

Derek Skillings

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date	Massimo Pigliucci
	Chair of Examining Committee
Date	Iakovos Vasiliou
	Executive Officer

Supervisory Committee:

Peter Godfrey-Smith

Catherine Wilson

David Papineau

Justin Garson

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Aspects of Biological Explanation
by
Derek Skillings
Advisor: Peter Godfrey-Smith

This dissertation is an evaluation of some strategies used for understanding the biological world. It argues that the complexity of living systems challenges the adequacy of traditional approaches to scientific explanation. An examination of the empirical details—especially those at the microscopic and nano scales—highlights the limitations of mechanistic explanation, common habits of causal reasoning, and theories of individuality. According to this analysis, starting with broad generalizations of how the world is, or a single universal theory of how it ought to be investigated or explained, has things entirely backwards. Instead, we ought to start by looking at the important processes, relations, and interactions uncovered by the sciences, carefully building towards general claims. In chapter one, I examine and challenge a dominant characterization of mechanistic explanation. I develop an alternative approach that situates biological processes as falling along a multidimensional gradient—with some processes being paradigmatic cases of mechanisms and some being marginal cases. In chapter two I trace recent developments in biology where the study of robustness has become increasingly important. I develop a general taxonomy of robustness to help clarify links between the study of robustness in different fields, and argue for a shift towards incorporating more top-down, system-level approaches into general accounts of explanation. In chapter three I examine aspects of causal reasoning in biology. I provide a novel distinction between two types of difficulties encountered while reasoning about biological systems. In the face of these problems I argue that approaches to explanation are best seen as heuristics, which focuses our attention on both how our conceptual tools resolve problems, and when and where they can break down. In the final chapter I turn to a debate at the interface of philosophy and biology about whether we need to reevaluate our understanding of the biological world because we—along with every other multicellular organism—are, and always have been, multispecies entities.

ACKNOWLEDGEMENTS

This dissertation is dedicated to Melissa and Persephone. I promise this is my last one…

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

i. Overview

Scientific explanation is one of the core issues in contemporary philosophy of science. Early philosophical accounts of explanation tend to emphasize conformance with laws and focus on the success of physics. The shortcomings of purely law-based accounts of explanation began to be realized when philosophers started paying attention to explanatory practice in biology. Much of the philosophical literature on explanation in biology falls within the causal and mechanistic traditions, and these approaches serve as both starting points and foils in my dissertation.

The explosion of progress in biophysics, molecular biology and genetics in the last few decades has painted a new picture of the inner workings of the living world and opened up new explanatory challenges. For example: How should we describe the causal relations of reliable processes that are driven by stochastic forces? What is the causal relation between a gene and a phenotype when the mechanism of gene expression is a robust network of pathways, few of which, if any, are independently necessary for the mechanism to function? How should we think of organisms given revelations about the ubiquity and interdependency of host-microbial interactions? In short, how should we explain the peculiar organizational and causal structures of the biological world?

My dissertation is organized as a series of philosophical problems arising from a close examination of the details of biological systems. A unifying theme of this project is a demonstration of how paying close attention to scientific details can transform philosophical problems. Each chapter focuses on particularly challenging aspects of biological systems, how

scientists have tackled these challenges when constructing theories or explanations, and the resulting problems posed for philosophical treatments of the relevant issue along with a strategy proposed to help overcome those problems. This order of analysis is intentional, as well as central to how I think the philosophy of science should be approached.

I maintain that starting with broad generalizations of how the world is, or a single theory of how it ought to be analyzed or investigated has the process entirely backwards. We shouldn't assume at the outset that there is an interesting unified category to be found. Instead we ought to start by looking at the important processes, relations, and interactions uncovered by the sciences to see if we can build up to generalities. Generality is a very important aim for science and philosophy of science. But I am convinced that we must carefully build towards general claims, rather than assuming them in the first place. Thus, the strategy that I pursue in my dissertation is to start by examining the details turned up by biological investigation, rather than offering from the start general claims that are supposed to hold for all biological explanation. We can then try to determine whether such general claims might emerge from the biological particulars.[1]

ii. Chapter Overviews

Biologists tend not to cite, nor search for, laws of nature in their explanations, and biological phenomena do not fall into regular, law-like patterns. "At least in biology, most scientists see their work as explaining types of phenomena by discovering mechanisms…" (Wimsatt 1972). In chapter one, I characterize a *basic account* of mechanistic explanation pulled from the philosophical literature. I present three challenges to this account using examples from molecular biology and biophysics. I argue that the *basic mechanistic account* is (1) insufficient

[1] See (Pradeu 2012; Guay and Pradeu 2016) for the advocation of a similar approach.

for explaining non-sequential and nonlinear dynamic processes, is (2) insufficient for explaining the inherently stochastic nature of many biological mechanisms, and (3) fails to give a proper framework for analyzing organization. I argue that biological processes are best approached as a multidimensional gradient—with some processes being paradigmatic cases of mechanisms and some being marginal cases. I then suggest three axes along which mechanisms vary within this gradient: isolability, organization, and sequentiality.

One striking feature of many biological systems is their high degree of robustness, or the ability of a system to maintain its function or behavior against perturbations and under a wide range of circumstances. Despite the ubiquity and importance of robustness, philosophers tend to think of it in a simple way. In chapter two I trace recent developments in two fields of biology where the study of robustness has become increasingly important: ecology and evo-devo. I argue for a shift towards incorporating more top-down, system-level approaches into accounts of explanation to better reflect scientific practice seen in those two fields. I develop a general taxonomy of robustness to help clarify links between the study of robustness in different fields. I define three general kinds of robustness: *redundancy, resistance and resilience*. Each of these categories are further subdivided in turn.

In chapter three I examine aspects of causal reasoning in biology. In my analysis, I provide a novel distinction between two types of difficulties encountered while reasoning about biological systems. First, there is a difficulty with imagining and predicting the behaviour of biological phenomena—especially across scales or domains--that stems from the stochastic and contingent nature of many biological processes. I refer to this as the *translation problem*. For example, reasoning about the emerging behavior of higher-level biological phenomena from lower-level phenomena can be difficult because the intracellular stochastically-driven biological

processes that make up the building blocks of life behave in a way that is alien to our everyday reasoning about the behavior of objects at our scale. Second, there is a difficulty with identifying all the numerous salient causal factors and interactions behind the production of biological phenomena. I call this the *interpretation problem*. It stems from the difficulty in decomposing the unique causal complexity that is found in robust, hierarchical and recursively organized biological systems[2]. In the face of these problems I argue that approaches to explanation are best seen as heuristics, which focuses our attention on both how our conceptual tools resolve problems, and when and where they can break down.

In the final chapter I turn to a debate at the interface of philosophy and biology about whether we need to reevaluate our understanding of the biological world because we—along with nearly every other multicellular organism—are, and always have been, multispecies entities. It is now widely accepted that microorganisms play many important roles in the lives of plants and animals. Every macroorganism has been shaped in some way by microorganisms. The recognition of the ubiquity and importance of microorganisms has led some to argue for a revolution in how we understand biological individuality and the primary units of natural selection. The term "holobiont" was introduced as a name for the biological unit made up by a host and all of its associated microorganisms, and much of this new debate about biological individuality has focused on whether holobionts are integrated individuals or communities. I show how parts of the holobiont can span both characterizations. I argue that most holobionts share more affinities with communities than they do with organisms, and that, except for maybe

[2] This problem, under various guises, has been noted and discussed by biologists for decades. For example, See Lewontin (1974, 1978) for extensive discussions about the epistemological difficulties in isolating and describing the evolutionarily significant genetic variation that explains change in a trait of interest.

in rare cases, holobionts do not meet the criteria for being organisms, evolutionary individuals, or units of selection.

The goal of the rest of this introduction is to briefly sketch out some general debates within the topic of scientific explanation in order to both situate and motivate the problems I take on in my dissertation.

iii. Three Contexts of Explanation

The term 'explanation' is used in three slightly different contexts in the philosophical literature, and accounts of explanation generally take one of these three as the fundamental context by which we can evaluate explanations. Sometimes, 'explanation' refers to a communicative act: explanations are a means of communicating scientific knowledge. Accounts that take this context as fundamental are pragmatic, and evaluate explanations by how well they convey understanding, (i.e., Bromberger 1965; van Fraassen 1980; Achinstein 1983). 'Explanation' might also refer to the facts about the world that are cited in an explanatory communicative act, such as the relevant laws, causes, and things out in the world. Accounts that take this context as fundamental evaluate explanations by whether or not they get the facts right, (i.e., Salmon 1970; Railton 1981; Salmon 1984; Lewis 1986; Woodward 2003; Craver 2007; Strevens 2008). Finally, 'explanation' might refer to the representational form that explanations take. Those that take representational form as fundamental to explanation evaluate explanations by whether or not they are adequate representations, (i.e., Hempel 1965; Kitcher 1981; Machamer et al. 2000). To a large extent, these three approaches to explanation—referred to respectively *as pragmatic, ontic,* and *representational*—are taken to be at odds with each other.

iv. A Brief History of the Philosophy of Scientific Explanation

The earliest account of something approaching explanation is usually agreed to be set out by Aristotle's account of causes in his Physics. Aristotle's account of causes set the agenda in Europe for the scholarly understanding of the natural world until the beginning of the 17th century (Schmaltz 2011). '[T]he search for causes of events in nature' that 'guided science from the time of Aristotle' was superseded at 'the dawn of modern science', starting with the work of Galileo, 'by a quest for laws of nature based on experiment and measurement' (Drake 1981: ix).

It was during this time in the early modern period that various mechanical philosophies began to be developed. Mechanical philosophy is usually associated with a reductive theory of matter. The parts of matter are defined by geometrical properties and move according to universal and mathematically-described laws of motion. All physical phenomena can be explained in terms of matter and motion. The mechanical philosophers had a problem in that the motions of corpuscular matter occur an unobservable level. Thus, scientific explanation consisted in forms of explanations where the geometrical properties and mathematical laws that hold of observable everyday phenomena, physics, are transferred, by analogy to those unobservable phenomena (Hattab 2001). The ultimate goal of scientific explanation was the mathematical demonstration of the phenomena in question (Hattab 2001, Schmaltz 2011).

The proponents of this kind of view of the world included Descartes, More, Leibniz, Boyle, Hooke, Gassendi, Digby, Hobbes, Locke, and more. Henry More was the first to use the term, but Robert Boyle popularized it, and Descartes is probably most closely associated with it, especially as it was applied to a new understanding of the workings of organisms (Hattab 2001). The early moderns, for the most part, rejected an Aristotelian conception of biological phenomena where generation, growth, nutrition, and locomotion are attributed to final causes,

and the activity of the vegetative and sensitive souls (Smith 2011) .

Descartes further pushed the application of mathematical and geometrical descriptions to nature when he identified nature as a divinely fashioned machine, stating that there is no difference in kind between the parts of nature and machines designed by humans, only a difference in degree (Hattab 2011, Smith 2011, Nicholson 2013). Living things are not merely like machines, they are really are just more complex machines. Descartes wrote in a letter to Morin: 'I know that you will say that the form of the clock is only an artificial form, while the form of the sun is natural and substantial; but I reply that this distinction concerns only the cause of these forms, and not at all their nature…'. He then claims that 'there is no essential difference' but only one of degree between them (translations quoted from Hattab 2011).

The idea of treating the working of organisms as analogous the workings of machines was not new to this period. Such a tradition goes all the way back to ancient Greece (Nicholson 2013). What was new was the conviction that it is not only useful, under certain circumstances, to treat organisms as if they are machines, but that we can only truly understand organisms if we conceive of them as machines (Nicholson 2013). A mechanical philosophy is thus able to bring together the animate and inanimate parts of the natural world (save the ensouled human), under one explanatory framework.

It is important to note that machines of the time were not just the well-known clock-work contraptions. The early moderns had a fairly sophisticated account of machines and organisms. Machines of that time also ran on air and water hydraulics, chemical and thermal processes. As characterized in a quote that I particularly like: "In short, a fermenting, exploding, end-directed machine of quasi-perpetual motion is a far cry from the sort of gear-driven contraption that

mechanical philosophers are supposed to have taken animals to be" (Smith 2011). But it was this complexity that also led to the demise, albight temporarily, of the mechanical framework for the understanding of organisms. A search for vital principles and vital forces dominated philosophical accounts of the organism until the 19th century, finally coming to an end in the early 20th century.

An approach to scientific explanation that also developed in the early modern period—and was perhaps ultimately more successful—was a shift to a view of the natural world based upon universal laws. The use of laws in early modern physics is, at least in part, explained by the increasing importance of 'mathematical principles' in the work of Descartes, Newton and others. Schmaltz (2011) traces the major moves in this history in the following way. Descartes, deviating from past scholastic practice, made laws central to his account of physics. Descartes did not contrast causes and laws, rather, he identified laws as the secondary causes of changes in motion that are ground in the internal features of bodies. The second major step was made by Malebranche, who took causality out of the material world and placed it in God. Causes were identified as God's efficacious volitions. The progression culminated in Berkeley's view (following Newton's work in physics) that the laws of physics are to be identified with mathematical principles that do not concern real causes of motion. At this point then, explanation in terms of scientific laws is a subsumption of particular events under inductive generalizations, as opposed to the illumination of causal structures (Schmaltz 2011). Scientific explanation characterized by the search for laws dominated until the mid-20th century when it came to be supplemented by a return to causal explanations.

The earliest contemporary philosophical accounts of explanation in science, such as (Hempel 1965), were one-size-fits-all approaches that tended to emphasize laws and focus on the

success of physics. These accounts were soon seen to be problematic, and later largely abandoned, for two reasons. The first reason came from actual explanatory practice in the special sciences. Exceptions to law-based accounts of explanation became common when philosophers started paying attention to fields such as biology. Biologists tend not to cite, nor search for, laws of nature in their explanations, and biological phenomena do not fall into regular, law-like patterns[3].

The second, and perhaps more decisive, reason for the abandonment of law-based accounts is that they aren't able to account for the apparent asymmetry of explanations. For example, given information about the angle of the sun and some basic knowledge of trigonometry one can derive the length of a flagpole's shadow from the height of the flagpole. Symmetrically, one can also derive the height of the flagpole from the length of the shadow. But it is the height of the flagpole that explains the length of the shadow, and not the length of the shadow that explains the height of the flagpole[4].

Following the increasing disfavor for law-based views of explanation, causation once again started to become central to most accounts of scientific explanation. Two questions have been historically central to understanding causation, and they still feature prominently in modern work on causal explanation. The first question is most commonly traced back to the works of J.S. Mill (Mill 1882). Among all of the factors that were necessary for the occurrence of some phenomena, which should be selectively highlighted as the cause/s in an explanation of that

---

[3] A notable exception is ecology. Ecologists search for laws, but have so far come to little agreement on whether there are any.

[4] The famous flagpole counter-example to Hempel's deductive model was introduced in Sylvain Bromberger's unpublished dissertation (Bromberger 1961). It was more widely disseminated through a more accessible description in Salmon (1971: 71-76).

phenomena, and which should be omitted or relegated to the level of background conditions? Call this the *selection problem*. The second question finds its roots in the works of David Hume. What is the relationship between explanations of singular events and explanations of regularities? Call this the *generalization problem*.

v. The Selection Problem

On very few accounts does the requirement to pick out the causes which brought about an event extend to all, or even very many, of the causal influences of that event. Namely, a good causal explanation only requires that we pick out the important causal influences. With these causal influences in hand there are two further types of causal selection that are often important in explanation: selecting the right 'level of explanation' and the prioritization of some causes through the omission of background conditions.

Selecting the right 'level of explanation' is often thought to be fairly straightforward.[5] The correct level of explanation is given by the level of what is to be explained. That level tends to be at the same level as the explanatory target, or one level down at the level of the parts or events making up the explanatory target. An explanation might be inadequate, or even fail as an explanation at all, if it is given at the wrong level given the explanatory target. For example, an explanation might be bad if it gives too many irrelevant low-level details. We don't--and currently couldn't--give an explanation of the rising of bread dough by giving micro-scale descriptions at the atomic level. The location and movement of every molecule, or even every yeast and wheat cell, is not necessary. It may even impede understanding. In most contexts we give explanations in terms of flour, water, yeast, salt and fermentation by-products.

[5] Being able to give a rigorous account of how this is to be done is much more elusive.

An explanation might also be inadequate because it is at too high of a level, such as when a mechanism is black-boxed. This level of explanation fails to reveal important explanatory details. For example, an account that just includes the fact that gas bubbles form in the dough and that those bubbles cause it to rise is inadequate. It doesn't explain how the bubbles form and makes no mention of yeast, fermentation, etc. Neither the too-high level nor the too-low level account are giving false information, they just appear to be missing the details that are relevant to the phenomena we are trying to explain.

When giving an explanation for the occurrence of an event we rarely cite all of the conditions, or causal factors, that were necessary for bringing about that event. Instead, we pick out a few, often just one, causal factors to give special priority, and relegate all the rest to the status of background or enabling conditions. We might say that it was the yeast that caused the bread dough to rise. Of course, we also know that, among other things, the yeast required water and nutrients, in the form of flour, to produce the gaseous byproducts that led to the rising of the bread dough. In contrast to selecting the right 'level of explanation', the decision about which causes to select as salient, and which causes to relegate through omission to 'background factors', is usually taken to be a purely pragmatic matter. This is the view that Mill took (1882: book 3, chapter 5). Many other prominent accounts of causal explanation take a similar stance in denying that there is any uninquely correct pragmatics of explanation (Railton 1981; Lewis 1986; Hall 2004; Strevens 2008).

vi. The Generality Problem

There are two types of targets for causal explanations: singular events and regularities. Likewise, there are two types of explanations corresponding to these two targets. Singular explanations are aimed at explaining singular events, such as the browning of a particular loaf of

bread in a particular oven at a particular time, and singular states of affairs, such as an oven being hot in a particular kitchen at a particular time. General explanations are aimed at explaining regularities, such as physical laws, theories, mechanisms or dispositional properties.

Claims about the relations between singular events and regularities usually stem from discussions about the nature of causation. For Hume, the truth of singular causal claims depends upon the existence of universal regularities. He writes: "[W]e may define a cause to be an object, followed by another, and where all the objects similar to the first, are followed by objects similar to the second". (Hume 1748: Sec. VII). We only say that throwing the rock caused the window to break, because of the regular association between rocks hitting windows and windows breaking. Regularity and causation are tied tightly together.

I will leave to the side questions about the order of precedence between singular causal claims and regularities. That is, whether singular causal events account for regularity, or whether we only attribute causation in singular events because of the observance of regularity. I will also be largely bracketing other questions about the metaphysics of causation. The discussion that follows is meant to apply no matter what the causal relation underlying causal explanation might be, e.g., regularity, Lewisian or interventionist counterfactuals, transference of a conserved quantity, probabilistic, etc.

vii. A Brief Survey of Causal Explanatory Accounts

Now that we have the problems that concern a causal account of scientific explanation on table, I will introduce some theories of causal explanation. This is not meant to be a comprehensive review. I offer a few accounts that best highlight the diversity of approaches in the literature.

vii.a Minimalist Accounts of Causal Explanation

The rise of contemporary causal approaches to explanation started with "minimalist accounts" like those of Railton (1981) and Salmon (1984). On Railton's account, the universe is a vast weave of causal influence, and for any given event it is the causes of that event which explain its occurrence. The explanatory targets are fundamental concrete events and every causal influencer must be cited in the explanation. An explanation requires all of the fine-grained details leading up to the target phenomena. In minimalist accounts, the gravitational pull of Saturn is a part of the explanation for the rising of bread dough. Given what we know about physical forces and the fact that gravity and electromagnetic forces never drop to absolute zero, everything in an event's past light-cone is a possible cause of that event. Railton uses the term 'ideal text' to refer to the full explanation of an event. Needless to say, we never have full explanations for events. The explanations that we give each other, and that scientists present in journals, are always partial. The selection of which causes we foreground in a partial explanation and at which level we explain is governed by pragmatics, convention, or norms of communication, not by any criteria for getting the explanation "right".

vii.b Difference-Maker Accounts of Causal Explanation

An account that requires the inclusion of every causal influence is neither useful nor satisfactory. It is, at best, a conceptual starting point for a richer theory of explanation. So we ask the question most commonly traced back to the works of J.S. Mill (1882): among all of the factors that were necessary for the occurrence of some phenomena, which should be selectively highlighted as the cause(s) in an explanation of that phenomena, and which should be omitted or relegated to the level of background conditions? More recent causal approaches to explanation tend to take only those causal influences that are difference-makers to be explanatory. To

understand a phenomenon is to see what made a difference to the causal production of the phenomenon and how it did so (Strevens 2008). Or as Jim Woodward (2010) puts it "…good explanations should both *include* information about all factors which are such that changes in them are associated with some change in the *explanandum-outcome* of interest and *not include* factors such that no changes in them are associated with changes in the *explanandum-outcome*."

I will distinguish between those causal influences that are difference-makers and those that are not by returning to my imaginary kitchen. While making a cake I drop an egg, it hits the hardwood floor and breaks. I want to explain why the egg broke. There are many causal influences on the breaking of the egg, among them are the hitting of the floor and the orientation of the egg as it leaves my hand. Only the floor makes a difference to the breaking, even though both affect exactly how the egg breaks. The orientation of the egg when it leaves my hand will influence where the egg first makes contact with the floor and how the force is transmitted through the egg. It will almost certainly affect the precise pattern of cracking and the trajectory of the splatter. However, these influences do not determine whether or not the egg breaks. They make a difference to the way that the egg breaks, but not to whether or not it breaks at all. Relative to the explanatory target of *whether or not the egg breaks*, the impact with the floor explains the breaking, the orientation of the egg on impact does not.

Difference-making accounts are aimed at higher-level events, rather than at fundamental concrete events. Fundamental concrete events include every single causal detail about an event. Any difference in the causal details, like a few molecules being switched around, makes for a different concrete event. The concrete event of the egg's breaking when I dropped it would not have occurred if anything at all had been different. If the orientation of the egg had been different, than a different concrete breaking event would have happened instead. This is because

all causal influencers are difference-makers for concrete events. Only when we abstract to the higher-level of that particular egg breaking at all do some causal influencers become difference makers.

There are a wide variety of criteria for picking out difference makers. Many of these criteria are the same ones used for picking out causal relations. There are counterfactual tests: a causal influence is a difference-maker for an event if, had the influence been absent, the event would not have occurred. There are probabilistic tests: a causal influence is a difference-maker only if the probability of the event occurring is higher when the influence is present than when it is not. I will look at one recent account for picking out difference-makers: Michael Streven's (2008) Kairetic account.

The kairetic procedure for building an explanation begins by building a model structured as a deductive derivation representing all of the causal influences that causally entail the event to be explained. Remove as many of the causal influences as you can from the model without invalidating the entailment. Next make the causal model as abstract as possible, by replacing the fine-grained details of the influences with more abstract descriptions of the same. Abstract to the point where any further abstraction would invalidate the entailment of the target event. You now have all the difference-makers for the event selected at the right level of explanation.

Strevens supplements the kairetic procedure with 'frameworks' in order to offer a complete account of explanatory practice. Something like a 'framework', or a *ceteris paribus* clause, is found in most difference-maker accounts. A framework specifies some set of conditions that is held fixed in an explanation. The framework is part of an explanatory request, and can be given either implicitly or explicitly. An explanation for an event that includes a

framework is not an explanation for that event in its totality, but rather an explanation for that event given whatever the framework specifies. For example, the request for an explanation of why the egg broke is not a request for why the egg broke—full stop—it is a request for why the egg broke given the effects of gravity, that we are not underwater, that the eggshell has not been artificially hardened, etc.

That gives us an answer to the *selection problem*, but what about the *generality problem*? The kairetic account shares the form of this answer with other difference-maker accounts of causal explanation. The explanation of a causal regularity and the explanation of any instance of that regularity invoke the same causal process. Strevens (2008) calls this the "first fundamental theorem of causal-mechanical explanation". The causal facts that explain any instance of a regularity can be used, in addition, to explain the regularity itself. In terms of the kairetic account, generalizations constitute a model schema for the explanation of any instance of that generalization. The explanation of an instance of a generalization comprises the explanation of the generalization together with the specifics of the instantiation, such as the filling in of variables within the schema with the particular values of the event. General explanations and singular explanations are produced by the same procedure; they are just at different levels of abstraction and have different explanatory targets.

vii.c Mechanistic Explanation

Explanation in the special sciences usually takes the form of isolating a regular phenomenon in nature and then providing the mechanism that accounts for that regularity (Wimsatt 1972). The most widely cited philosophical account of mechanisms defines them as follows: "Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions" (Machamer et al. 2000).

The first feature of mechanisms is that they are hierarchically structured and can be broken down into components. The second important feature of a mechanism is that the entities in a mechanism must be appropriately located, structured, and oriented, and the activities in which they engage must have a temporal order, rate, and duration (Bechtel and Richardson 1993; Machamer et al. 2000). The organization of entities and activities determines the ways in which they are able to co-produce phenomena. Finally, the entities and activities constituting the mechanism must work in a regular fashion. A mechanism regularly produces or gives rise to the same phenomenon because earlier stages lead reliably to final stages, so that different instances of a mechanism share patterns of activity among similar or identical entities (Andersen 2012).

Mechanistic explanations are primarily used to pick out the structure of processes that are responsible for producing higher-level regularities. In many ways mechanistic explanation is a refinement of the causal explanation that we saw above set to a particular task, the explanation of regularities. Mechanistic explanations tend to focus on the structure of the whole causal process rather than the nature of the relations between individual events in the process. Though it might not be necessary to underlie a mechanistic account with a difference-maker account, most seem to have one in mind.

viii. Towards a More Nuanced Account of Explanation

The survey of explanatory accounts that I just presented reads as a build-up of explanatory desiderata. This is intentional, and to some extent tracks the 20th century history of philosophical accounts of explanation. The transition between causal explanatory "genres" rests upon a debate about what a complete explanation looks like in an ideal case. For the minimalist account, a complete causal explanation consists of all the microphysical details of causal influencers of the event to be explained. Someone subscribing to a difference-maker won't be

satisfied with the minimalist account. It's not that the minimalist explanation is wrong about the facts, it's just that it is missing some information. In addition to knowing all of the causal influencers, this person also wants to know which of these influencers are difference-makers in regards to the explanandum. The minimalist explanation needs to be supplemented in order for it to be an adequate explanation. If we have all the causal facts that are contained in a minimalist explanation, then we have all the causal facts necessary to generate a difference-maker explanation. An adequate explanation requires that we make difference-making causal influences explicit.

I see a similar tension between the difference-maker accounts and the mechanistic accounts. Again, it's not that a difference-maker explanation is wrong about the facts, it's just that it is missing some information. The mechanist is interested in explaining regularities. In addition to knowing about the difference-making parts in the causal chain leading to the explanandum, the mechanist wants to know which structural relations in that causal chain are important for reliably producing the explanandum[6]. When we are trying to explain regular phenomena that are produced by causal processes, the difference-maker explanation needs to be supplemented with mechanisms details in order to be considered explanatorily adequate.

It is almost certainly not the case that we will always be able to build a complete mechanistic explanation out of a complete explanation made up only of difference-makers. This is because a complete mechanistic explanation requires that we make the structure of the mechanism explicit in addition to picking out the difference-makers. In some cases, that

[6] This line of argument I take to be similar to the one used to justify *asymptotic explanations* by Batterman (2002).

information might be retrievable from the complete difference-maker explanation. In cases where the mechanism contains redundant or back-up processes, these non-difference makers might have been left out of the difference-maker explanation.

I will be pushing this explanatory expansionist line further, and doing so guided by the biology of systems that have received very little attention in philosophy. One important consequence of my analysis is that it reveals how biologists have adapted their explanatory practices, and shows how explanation is more diverse than mainstream philosophical accounts suggest. Paying close attention to the behavior, complexity, and organization of causal processes in living systems, and their differences from other systems, motivates further revisions to the received views of scientific explanation. The available mechanistic and difference-maker accounts of explanation have enriched our explanatory reach, but they fall short in their goal of adequately explaining all types of causal phenomena. Guided by the biology—by the details of how living systems actually work—I show how important features like stochasticity, robustness and recursive hierarchical organization are not accommodated by dominant accounts of explanation.

A final comment in anticipation of responses to what I have said in this section. The transition between causal explanatory accounts through the addition of explanatory criteria is supposed to be motivated by finding the more minimal account unsatisfactory. When faced with an argument for the centrality of difference-makers, the minimalist might counter that, while they can see the pragmatic value of highlighting difference-makers, they see no reason to think that only the difference-makers are explanatory. After all, the difference-makers are included in the set of causal influences, so the minimalist explanation isn't missing anything at all. Similarly, when faced with an argument for the importance of picking out mechanisms over-and-above the

difference-makers, the difference-maker adherent might counter by claiming that the additional explanatory requirement is redundant once you have the set of causal processes that instantiate the purported mechanism. I expect a similar reply could be made in response to the addition of the explanatory criteria I will suggest over the course of my dissertation.

I don't think there are any decisive arguments against these replies. The defender of the more extended account might try to further motivate the critic by an appeal to usefulness, predictive power, or a shared search for the most explicit account of our explanatory practice. There are good reasons to value the more extended accounts that are based on actual explanatory practice in science. There is also a long tradition—and a strong intuition—that says the gravitational pull of the distant stars is explanatory irrelevant to the explanation of the breaking of an egg. Beyond this, the most that someone offering a more extensive list of explanatory requirements might hope for is to have others see, as they do, the explanatory force behind their answers to questions about why things are the way they are.

With this thought in mind, I can now begin my project, showing how the details of the biological world motivates an expansion in our approaches to explanation, a revision of our everyday causal reasoning habits, and an appreciation of ourselves as communities.

CHAPTER 1: Mechanistic Explanation of Biological Processes[7]

1.1 Introduction

The biological world is dynamic and evolving. One of its most prominent features is change, whether that change is an oscillating cell in homeostatic equilibrium, a developing multicellular organism, an evolving lineage, or seasonal ecosystem cycling. For much of biology providing a satisfactory explanation for a particular change means providing the *mechanism* that accounts for the phenomenon in question. A substantial amount of recent work in the philosophy of biology has been directed towards offering analyses of mechanisms and mechanistic explanation. Levy (2013) argues that this explosion of activity has led to mechanistic talk being adopted by three closely related, but often not well distinguished, projects (see Nicholson 2012 for a historical division of mechanistic views). The first project provides a mechanistic account of causation. It is the view that causal relations between non-fundamental physical phenomena exist in virtue of underlying mechanisms. The second project is concerned with explanation. A mechanistic view of explanation holds that certain types of phenomena, e.g. biological phenomena, are explained by identifying the mechanisms generating those phenomena, i.e. by specifying the underlying parts, their organization and their interactions. The third project is about how phenomena are best modeled. Levy (2013) uses the label "strategic mechanism" for the thesis that certain phenomena are most readily understood when modeled abstractly as mechanisms. This chapter is concerned with the mechanisms—and their underlying components--identified by the explanations in the second project, as well as the target systems to be modeled

[7] This chapter was published in a slightly modified form as: Skillings DJ. 2015. Mechanistic Explanation of Biological Processes. Philosophy of Science. 82:1139-1151.

in a type three project. That is, I will primarily be focused on the entities underlying a phenomenon rather than the representation of those entities in a model of that phenomenon. I will leave to the side the discussion of the nature of causal relations, as it is, at best, tangential to my goals.

In this paper I will characterize a basic account of mechanistic explanation by presenting a recent and prominent approach in the philosophy of biology, almost universally referred to as the "new mechanist" approach (Section 2). I will highlight tensions between this basic approach and approaches that emphasize the non-sequential and stochastic character of some biological phenomena. Examples from recent research in molecular biology will be used to illustrate this point (Section 3). I will suggest that the best way to categorize the diversity seen across putative mechanisms is to view it as a multi-dimensional gradient--with some entities being clear cases of paradigm mechanisms and some entities being marginal cases--as opposed to being made up of discrete categories. I propose that this is accomplished by situating processes of interest within a multi-dimensional space generated by the interaction of three features: isolability, organization and sequentiality (Section 4). My overall goal is not to dispense entirely with the basic new mechanist approach, but rather to find ways to extend the account so that it is adequate for handling the full diversity of biological processes that are putative mechanisms.

1.2 Mechanisms

The most widely cited philosophical account of mechanisms defines them as follows: "Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions" (Machamer et al. 2000, hereafter MDC). Other prominent accounts also define mechanisms by decomposing them in terms of parts and their activities or interactions: "…a mechanism is construed as generating a

phenomenon through a start-to-finish sequence of qualitatively characterized operations performed by component parts" (Bechtel and Richardson 1993; Bechtel 2011); "A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations" (Glennan 1996, 2002).

These three accounts serve as core examples of what is commonly called the *new mechanist* approach. Glennan focuses on parts, their properties, and their interactions. An interaction is just a change in the property of one or more parts that *brings about* a change in the property of another part. The ability to produce change is a feature of entities. MDC have an expanded ontology: entities don't just interact, they engage in activities. Activities exist independently of entities on this account (Machamer et al. 2000; Glennan 2002; Illari and Williamson 2012, 2013). Activities are the producers of change, and entities are the things (with their properties) that engage in activities (Machamer et al. 2000). An important feature of activities is that they only exist extended in time, unlike entities, which can be extended in time but need not be. Bechtel and Richardson (1993) emphasize how scientists decompose mechanisms structurally into their parts and functionally into their operations. They emphasize the same distinctions as MDC (entities/parts and activities/operations) but prefer the terminology of parts operating on each in order to highlight the direct involvement and multiple roles of parts, and are wary of making any metaphysical commitments. With these differences in mind, I will use the terms *basic mechanism* and *basic mechanistic account* to refer to a mechanism as outlined by the MDC account.

It will be helpful at this point to highlight with examples the features of a basic mechanism. The first feature of mechanisms is that they are hierarchically structured and can be

broken down into components. The second important feature of a basic mechanism is that the entities in a mechanism must be appropriately located, structured, and oriented, and the activities in which they engage must have a temporal order, rate, and duration (Machamer et al. 2000). Activities connect entities into a coherent process. The organization of entities and activities determines the ways in which they are able to co-produce phenomena. For example, a neurotransmitter (entity) and a receptor (entity) bind (an activity) by virtue of their structural properties and charge distributions. Similarly, a DNA base (entity) and a complementary DNA base (entity) hydrogen-bond (activity) because of their geometric structures and weak charges. Finally, the entities and activities constituting the mechanism must work in a regular fashion. A mechanism regularly produces or gives rise to the same phenomenon because earlier stages lead reliably to final stages, so that different instances of a mechanism share patterns of activity among similar or identical entities (Andersen 2012). What makes a mechanism regular is the productive continuity between stages (Machamer et al. 2000). Discontinuities are one way to mark the bounds of mechanisms[8].

An example of a basic mechanism comes from the text book description of the mechanism for protein synthesis in prokaryotes. At its simplest, the protein synthesis mechanism is a process in which many small entities, amino acids, are joined together by a ribosome to form a larger entity, a protein [figure 1.1 and the description that follows based on Moore (2012)]. The initiation of protein synthesis begins with the recruitment of ribosomal subunits from the cellular pool by initiation factors. The small ribosomal subunit then binds to an appropriately aligned mRNA and an fMet tRNA. Initiation ends when the second amino acid in the protein sequence is

[8] Discontinuities can themselves be thought of in terms of strong modularity, in the (Woodward 2003) sense of modularity. See Section 1.3.3 for a further discussion of modularity.

carried to the ribosomal complex by a tRNA. Elongation of the polypeptide then ensues in a cyclical fashion. During elongation, a tRNA carrying an amino acid binds to the ribosome in what is called the A site. The ribosomal complex moves the tRNA from the A site into the P site, and another tRNA carrying the n+1 amino acid moves into the A site. Within the P site the ribosome then transfers the amino acid from the tRNA onto the end of the growing protein in a process called peptidyl transfer. At this point the ribosome must be reinitialized before the next amino acid can be added to the nascent protein. During this process, which is called translocation, the amino acid-less tRNA in the P site leaves the ribosome, the n+1 tRNA in the A site moves to the P site, and the ribosome moves along its mRNA in the 3' direction by one codon. When the translating ribosome encounters a stop codon in the mRNA, the protein is released from the ribosome and the two ribosomal subunits return to the cellular pool. This process is called termination.

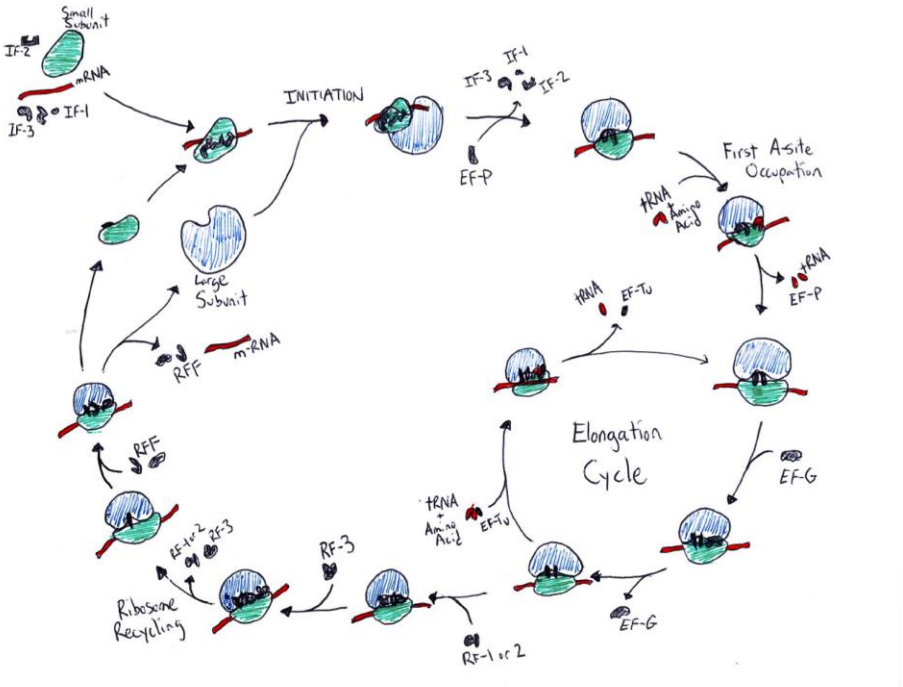

Figure 1.1 Protein synthesis cycle based on illustration found in Moore (2012).

1.3 Criticisms of the Basic Mechanistic Account

Biological explanations that adhere to a basic mechanistic account like the one above have often been criticized as inadequately representing and explaining many biological phenomena. I will group these criticisms into three types. First, treating mechanisms as paradigmatically sequential and linear fails to accommodate a vast number of dynamic biological processes that are non-sequential, oscillatory, branching or feedback systems (Bechtel and Abrahamsen 2005, 2013; Bechtel 2011). Second, a strong requirement for regularity in mechanisms fails to accommodate the biological processes that are internally and temporally stochastic as a result of their dependence on the levels of free energy in the system. Finally, the basic mechanistic accounts fail to adequately distinguish between processes that are organized in an orderly way with parts that have differential roles and local relations, and those processes in which such features are absent (Levy and Bechtel 2013; Levy 2014).

1.3.1 Non-Sequential Dynamic Mechanisms

The basic mechanistic account emphasizes that a mechanism moves from start or set-up conditions to finish or termination conditions in an ordered and sequential way (Machamer et al. 2000). In a series of papers, William Bechtel and Adele Abrahamsen argue that the basic mechanistic account is limited by its emphasis on sequences of operations (Bechtel and Abrahamsen 2005, 2013; Bechtel 2011, 2012, 2013). They advance the term *dynamic mechanistic explanation* for an extended account that includes non-sequentially organized mechanisms.

The most straightforward and tractable organization that a process can have is one in which there is spatial adjacency and connectedness in its parts, and a stepwise temporal sequence

in its operations. Machine-like mechanisms are often organized in this way. More complex sequential organizations are also common. The Krebs cycle is not a start to finish sequence, but rather a sequence that runs in a loop. Many biochemical mechanisms are also built around some sort of negative feedback loop, a process in which the result of the process regulates the process itself.

Understanding these more complex sequences requires one to become increasingly more concerned with the dynamics of mechanisms, or how changes in the environment can affect the rate, duration and activation of a mechanism. The basic mechanistic account can accommodate this low-level dynamical behavior as long as it is sequential. But many biological mechanisms exhibit oscillations or more complex dynamical behavior, and this can be crucial for orchestrating operations within the mechanism (Bechtel and Abrahamsen 2013). A challenge to the basic mechanistic account comes from complex dynamical behavior that emerges when mechanism organization is *non-sequential.* Organization might be non-sequential because there are forks that can take the mechanism in different directions. Non-sequential organization might also arise from redundant but different sub-mechanisms, such that final product or state of the primary mechanism is unaffected, but the path to that point can differ. Given a particular end-state it might be impossible to determine the specific path a non-sequential mechanism traveled to get there. This is not the case for a sequential mechanism. This doesn't mean that a non-sequential mechanism doesn't produce regular changes, or that the parts and relations of the mechanism are indescribable. Rather, it is a point about how in any given instance of the mechanism, the unfolding activity of that mechanism might differ in response to differences in start-up conditions or ongoing environmental inputs.

1.3.2 Stochastic Mechanisms

Under the basic mechanistic account, a process must behave regularly in order for it to be considered a mechanism. What makes a mechanism regular is the productive continuity between stages (Machamer et al. 2000). A mechanism regularly produces or gives rise to the same phenomenon because earlier stages lead reliably to final stages, so that different instances of a mechanism share patterns of activity among similar or identical entities (Andersen 2012). A strong requirement for regularity in mechanisms fails to accommodate the biological processes that are internally and temporally stochastic**,** such as when their mode of operation is highly dependent on the levels of free energy in the system. In these types of systems, changes are probabilistic rather than regular. In many cases, instances of a stochastic mechanism do not share the exact same pattern of activity among even identical entities.

Let's return to the example of protein synthesis. The mechanism outlined in Section 1 was a version of the classic textbook explanation of protein synthesis. The ribosome is one of the primary players in the mechanism of protein synthesis. As more has been learned about it, the picture of protein synthesis has begun to change. Moore (2012) characterizes the progress in describing the ribosome since the first atomic resolution crystal structures of ribosomal subunits were published in 2000. The explicit goal of many structural biochemists was and is to provide a "movie" of protein synthesis. In this project the ribosome is characterized at discrete stages. The protein synthesis movie is created by characterizing all the different states of the ribosome and how they are related to each other, these are still frames of the movie that are put in order and played to show the process of protein synthesis. At the heart of the movie is a ribosome that is doing something. There are regular changes between stages, and earlier stages lead reliably to later stages.

Moore argues that these structure-based movies of the ribosome make it appear to be something that it is not. The reason that the movies are misleading is because ribosomes don't behave like machines that can be explained mechanically[9]. Moore states that: "[A]ll the functionally significant movements of the ribosome, both internal and external, are biased random walks, and it is most unlikely that any given ribosome will ever do exactly the same thing twice as it elongates some polypeptide" and "It follows that if a movie of elongation is not to be totally misleading, it must depict the endless series of meaningless, thermally driven, conformational fluctuations that separate one functionally significant event from the next." He predicts that 99.99% of the activity in protein elongation is random fluctuation that has nothing to do with function and that most of the progress through the protein elongation cycle is driven by changes in relative free energies coming from conformational state changes.

Think of a machine like a mechanical clock. The workings of the mechanical clock are explained by looking at how the physical parts of its mechanism interact. The mainspring stores mechanical energy through torsion. As the mainspring unwinds it rotates a central gear, which in turn drives the rest of the gear train. The last gear engages the escapement. The escapement is also engaged by the balance wheel, a weighted wheel which keeps time by rotating back and forth at fixed intervals according to the tuning of a balance spring which pushes back on it. The interaction between the mainspring and the balance wheel via the escapement mechanism turn the gears at a fixed rate. These gears are attached to the rotating hands of the clock face, which indicate the time. The movement and interactions of the parts of the watch explain how the watch works.

---

[9] See Pigliucci and Boudry (2011) and Boudry and Pigliucci (2013) for further discussion of some problems with machine metaphors in biology.

The parts of a protein, like a ribosome, don't stand in the same relations as the parts of a mechanical clock. The mainspring of a watch can reliably turn the gear train. Macromolecules immersed in liquids are at the mercy of frictional and thermodynamic forces. At any given time, the forces produced by a ribosome are one-tenth of the level or less than the average thermodynamic forces experienced by that ribosome. This means that the ribosome is randomly pushed into different conformational states by external forces, rather than moving in sequence according to the activity and interactions of its pars. Thus, all the functionally significant activities of the ribosome, both internal and external, in protein synthesis are probabilistic (Moore 2012). The probability that a ribosome will engage in an activity that is important for protein synthesis depends on the free energy in the system and the presence of protein factors that promote some states over others.

Moore recommends that we "come to think of the ribosome as the dynamic, constantly varying structure it is and always was" and move away from quasi-mechanical explanations for the properties of ribosomes and "instead seek to understand them in terms of the particle's conformational energy landscape." He boldly claims that structure-based movies of macromolecular processes have no explanatory power.

The mechanism of protein synthesis doesn't neatly fit into the basic mechanistic account. There is no productive continuity between stages, where earlier stages directly produce later stages. Rather, the mechanism proceeds stochastically. Movement through the process of protein synthesis is a biased random walk. Watches reliably tell time, and ribosomes reliably produce proteins, but the relations between the temporal stages of the mechanisms are much different. A watch has a fixed structure and operates in a fixed sequence. The parts of the ribosome are constantly in motion and changing their relations to each other due to forces that dominate at the

molecular level, such as: atomic bonds, friction, thermal forces and internal vibrational decay. These ongoing interactions are what dynamically stabilize the ribosome even as it flows constantly from conformation to conformation in a biased random walk. Furthermore, it is this constant change that leads to the functionally important states needed for protein synthesis. The ribosome is a dynamic, constantly varying entity, and it is this feature that is most explanatory when it comes to understanding how the ribosome synthesizes proteins.

There is another notion of stochasticity in the context of mechanisms that is discussed in Glennan (1997). Glennan is primarily concerned with causation and the relation between stochastic and deterministic processes. On Glennan's account of causation (1996, 1997), "two events are causally related if and only if those events are connected by a mechanism". A mechanism, in turn, is "a complex system consisting of a number of parts which interact to produce some outward behavior" (Glennan 1997: 509). The mechanism has "a set of observable dispositions" that are explained by the interactions of the mechanism's parts, which have their own dispositions (Glennan 1997). In regards to stochasticity, the relation between the dispositions of a mechanism and the dispositions of its parts can be related in three ways: stochastic/deterministic, stochastic/stochastic, and deterministic/stochastic. Parts with stochastic dispositions can underlie a deterministic mechanism. For example, the random behavior of individual gas molecules leads to the deterministic behavior of gases at the macro level as described by Boyle's law. Likewise, chaotic systems where a small change in initial conditions can lead to vastly different outcomes are (effectively) stochastic even if the underlying parts all behave in deterministic ways. It is important to note that Glennan's account is primarily

concerned with our subjective understanding of processes[10]. It doesn't matter if the process is actually objectively random or deterministic, it only matters if it is effectively stochastic or deterministic from our point of view.

An important distinction between my view and Glennan's is that Glennan is giving a theory of causation by which causality is captured in terms of the interactions of physical processes. It doesn't matter if the interactions are stochastic or deterministic. My focus is on a much more restricted set of phenomena, those hierarchically structured processes that are organized (to some extent) and show show regularity (to some extent). On my view, not every causal process ought to be, by definition, mechanistic. Mechanisms feature in causal explanations, but the metaphysical details need not be restricted to Glennan's mechanistic theory of causation. My account of mechanistic explanation is consistent with Glennan's mechanistic theory of causation, but not limited to it.

1.3.3 Order in Mechanisms

The basic mechanistic account requires that the entities in a process are organized if they are to be considered a mechanism: parts must be appropriately located, structured, and oriented, and the activities in which they engage must have a temporal order, rate, and duration. This account fails to adequately distinguish between processes that are organized in an orderly way with parts that have differential roles and local relations, and those processes in which such

[10] Howick (2011) also talks about stochastic mechanisms in the context of the debate between evidence-based medicine versus comparative clinical studies. Howick discusses how mechanisms often have stochastic behaviors, but this has to do with our ability to predict outcomes. The apparent stochasticity is thus subjective and may be the result of our incomplete knowledge about the mechanism. This differs from my notion of stochasticity, where the randomness is a basic feature of the mechanism under consideration. I thank Justin Garson for bringing the Glennan and Howick discussions of stochasticity to my attention.

features are absent (Levy and Bechtel 2013; Levy 2014).

Consider two processes in the mechanism of Ca^{2+} regulation: the passive diffusion of Ca^{2+} from the interior of the mitochondria out into the cytosol, and the active channeling of Ca^{2+} across a membrane. In the case of diffusion there is a system that consists of a semi-permeable membrane that straddles an electrochemical gradient. The exterior environment is more negatively charged than the interior environment and the positive Ca^{2+} ions will move across the membrane until electrostatic equilibrium is reached. The rate of diffusion is primarily a factor of the permeability of the membrane and the difference in charge between the two environments.

In the case of active channeling, when Ca^{2+} levels in the cytoplasm drop, secondary messaging systems within the cell can activate protein complexes that release Ca^{2+} stores from inside the endoplasmic reticulum. This example describes ryanodine receptor 2 (RYR2). RYR2 is a membrane-bound complex composed of four subunits that form a channel for Ca^{2+} to cross (figure 2). The complex is associated with various proteins that function to modulate its opening including: factors (like CSQ) that modulate the sensitivity of RYR2, transmembrane proteins that facilitate the opening, multiple enzymes that bind and control the "powering" of RYR2 by cAMP, and scaffolding proteins that bind further enzymes that reset and close the channel.

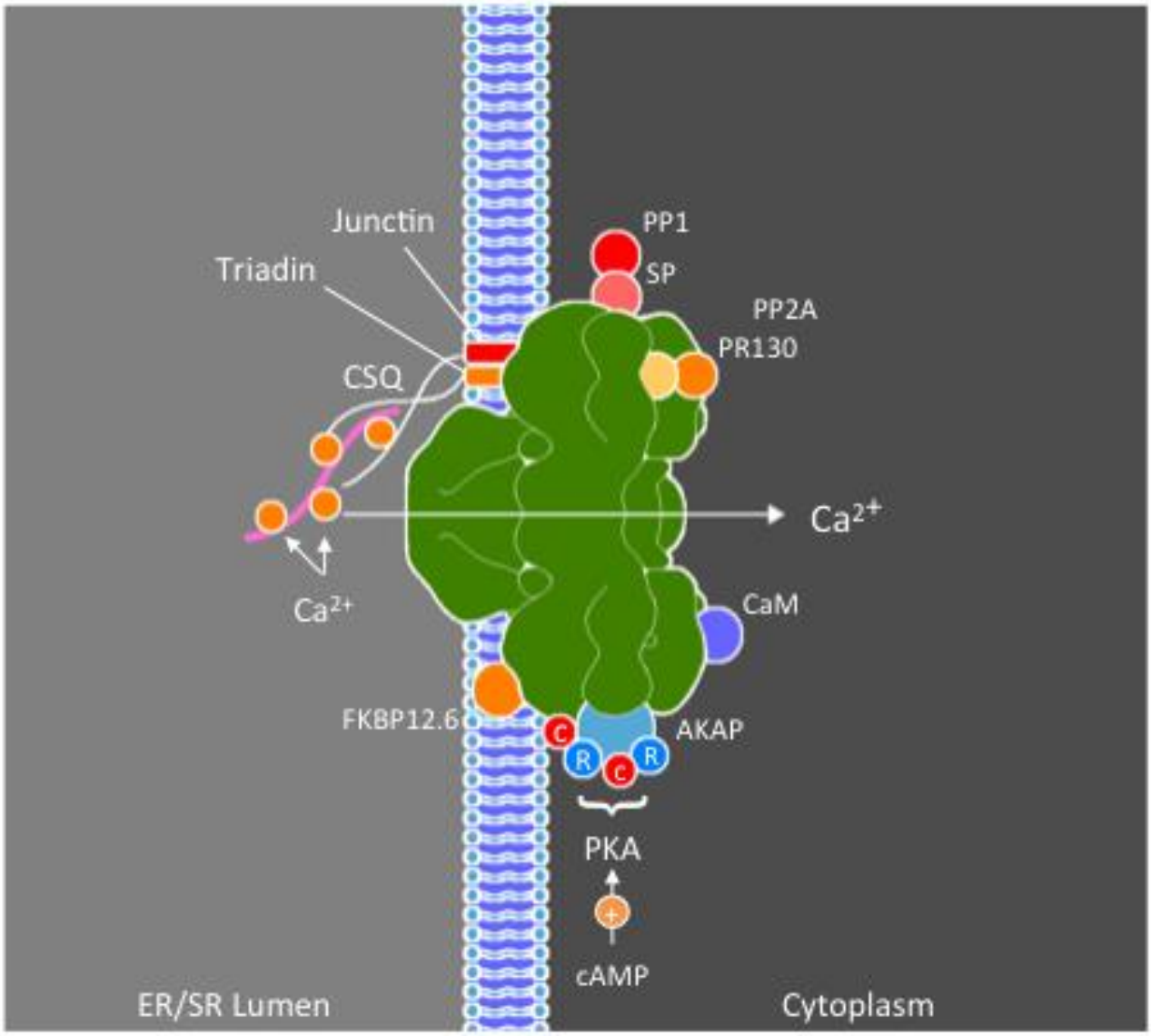

Figure 1.2 Ryanodine receptor 2 (RYR2) based on Berridge et al. (2003). Illustration by Shawn Simpson

Both of these example processes are dependent on system-level organization and feature interactions between constituent parts. One important difference is in how they are ordered. Active channeling is a highly ordered phenomenon: every component must be in the right place, at the right time, playing the right role. Diffusion, on the other hand, is a disorderly phenomenon: the role and layout of individual particles is insignificant. It doesn't matter what ions help generate the electrochemical difference across the membrane from the high Ca^{2+} concentration,

nor does it matter what the composition of the membrane is as long as it is semi-permeable. All that matters are the average properties of the components of the system.

Levy (2014) gives a characterization of orderliness that I will adopt here. Suppose we have a system S, exhibiting a behavior B. S is orderly to the extent that:

(a) Distinct components of S play different roles in bringing about B.

(b) Components play their roles in virtue of local relations to other components.

An orderly system exhibits an internal division of labor, analogous to that present in many manmade machines: each part does something distinct and recognizable, but there is also interdependence among parts, so that the system's overall behavior is an integrated product of their activities. Component parts can be distinct in two ways. They might be distinct because they are spatially separated. They might also be functionally distinct, i.e., they are distinct with respect to the difference made by the components. Levy fleshes this out in terms of Woodward's conception of modularity (Levy 2014; Woodward 2003). A component's contribution is modular if the difference it makes is independent of the difference made by other components. That is: component X's contribution is modular if it is possible to disrupt the activity of other components without affecting the contribution of X. In RYR2 channeling, for instance, it may be possible to deform the transmembrane proteins that help stabilize the size of the channel opening, thus disrupting channel size, without affecting the components that flip the RYR2 complex between open and closed states, and vice versa. These would then be distinct components of the active channeling mechanism. This functional distinctness in roles is what contributes to higher levels of orderliness.

The requirement that components play their roles in virtue of local relations to other components is related to the idea that an orderly system exhibits internal integration. It is possible for a system to have functionally distinct parts, i.e. for its parts to contribute differentially, without those parts being integrated. For example, the electrochemical charge of the cell is determined by the mixture of positive and negative ions within the cytoplasm. There are many different components that play different roles, but they do so by freely mixing, such that the specific layout of elements does not matter much. It is an aggregate property of the system (Wimsatt 1986). This is different than an active membrane channel system, there the local relations between parts are very important for the proper functioning of the process. In regards to electrochemical charge, it won't matter much if you switch two Ca^{2+} ions on opposite sides of cell with each other, or if you swap a Ca^{2+} for two H^{+} or a Mg^{2+}. In contrast, an active membrane channel will be greatly disrupted if a transmembrane protein is swapped with a receptor that binds cAMP.

1.4 Dealing with Mechanistic Diversity

I have given a basic account of mechanistic diversity and then categorized three ways in which this basic account can be found to be lacking. None of the criticisms I have discussed have challenged the existence of mechanistic phenomena, but rather find the basic account inadequate for handling the diversity of processes that are putative mechanisms. One route to dealing with this diversity of mechanism-like phenomena is to try to and isolate *the* true mechanisms by giving a more substantial theory. Another route is to subdivide mechanisms into smaller discrete categories: machine mechanisms, biological mechanisms, dynamic mechanisms, etc. A third approach begins by recognizing that there is gradient from phenomena that are produced by processes that are paradigmatically machine-like in their mechanistic functioning, to phenomena

that are produced by processes that don't conform to even the most minimal and basic account of mechanisms. The approach here is not to try and divide this space up discretely, but to describe the dimensions along which these phenomena grade into each other. I offer a brief account of mechanisms in the spirit of this third approach following the example set by the treatment of biological populations in Godfrey-Smith (2009).

Once again, my starting point is the basic mechanistic account. The minimal goal in mechanistic explanation is to explain a phenomenon by identifying the lower-level phenomena that produce the higher-level target phenomena. This requires identifying the lower-level entities and characterizing what they do and how they interact with other. This is the minimal notion I will use for what is required to consider something as even a putative mechanism: a hierarchically structured system where the higher-level can be at least partially decomposed into lower-level parts. Not all phenomena with these characteristics will be mechanisms, but it is possible to give the features of systems which distinguish the clear or paradigm cases of mechanisms from more marginal ones. Three interacting parameters will be used to make this distinction (figure 1.3).

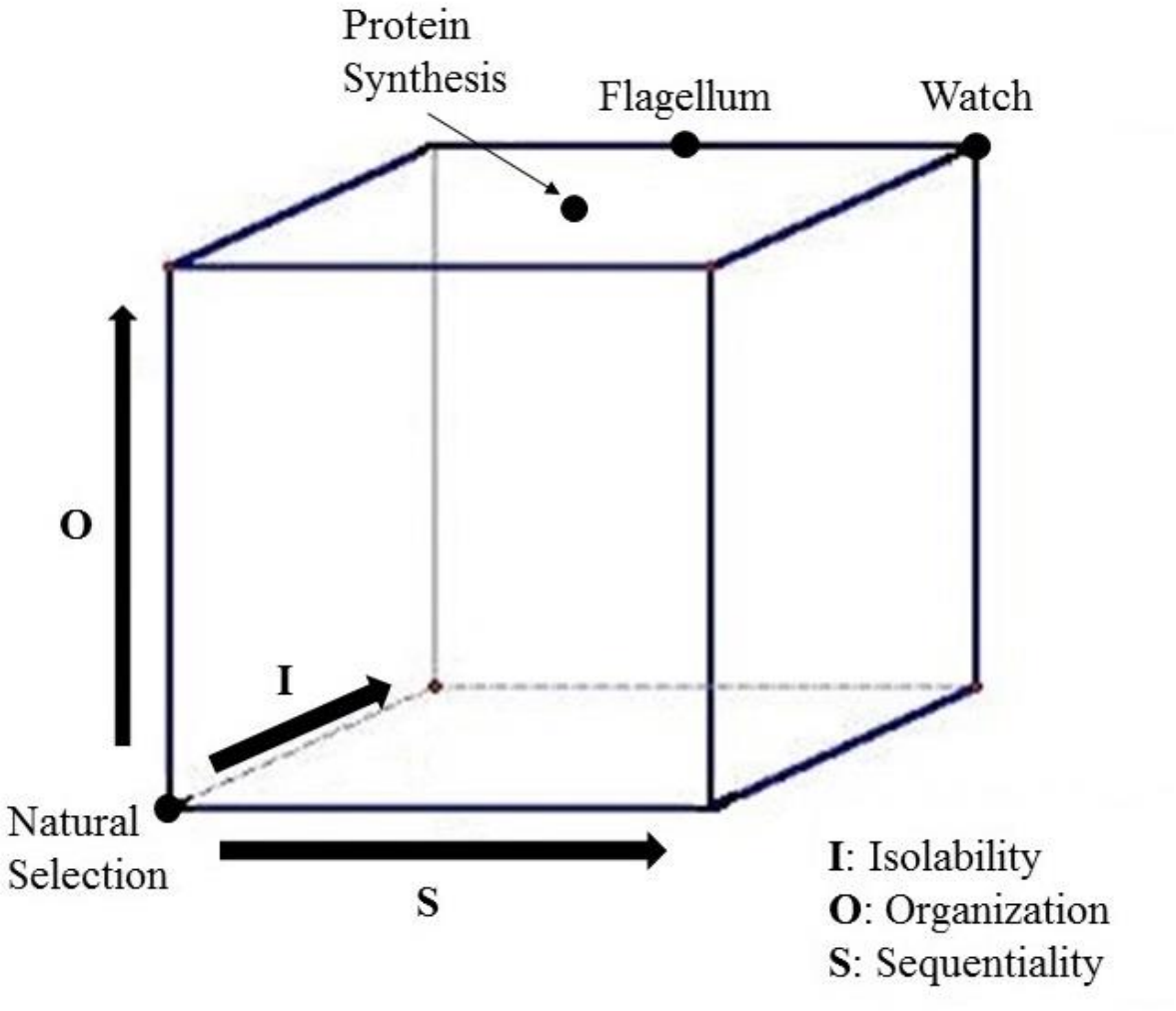

Figure 1.3 Three dimensions of mechanism-related processes.

The first parameter is the isolability of the system. This is the measure of the overall sharpness of the boundary between the system and its environment. Identifying the start-up, maintenance and finishing conditions of a process requires being able to identify what is part of the process and what is input from, or output to, the environment. An example of a system with sharp boundaries is a mechanical watch. An example of a system with vague boundaries is a population undergoing speciation with geneflow. An intermediate case might be synapse firing.

The second parameter is the organization of the system. This is the measure of the

orderliness of the system in terms of the distinctiveness of its component parts, the discreteness of its component parts, and the importance of local relations between parts in producing the higher-level phenomena. Highly ordered systems include bacterial flagella, and again, a mechanical watch. A system with low order is two liquids undergoing passive diffusion across a membrane, and a system with intermediate order might be a population undergoing selection.

The third parameter is the sequentiality of the system. This is a measure of the linearity and stochasticity of the system. A system with low sequentiality might be characterized by feedback loops, cycling, oscillations, variation in how component parts interact, or stochastic activities. A system with low sequentiality is the process of Ca^{2+} homeostasis. A moderately sequential system is the process of protein synthesis. A highly sequential system is a coin-operated gumball machine.

The clearest cases of mechanisms are machine mechanisms like a mechanical watch, which have a high degree of isolability, organization and sequentiality. These are the systems that most clearly meet the criteria of the basic mechanistic account. An example that is at best a marginal case of a mechanism is evolution by natural selection, a process that has a lower degree of isolability, organization and sequentiality. Many biological processes that are taken to be true mechanisms are highly isolable and ordered, but are only moderately sequential. There are a large number of intermediate cases.

There is an increasing recognition of the importance of cyclical, oscillating and biased stochastic processes for the explanation of biological phenomena. Using this multi-dimensional framework could help to extend the basic account of mechanistic explanation so that it can deal with biological mechanisms that are non-sequential, stochastic or less-orderly.

CHAPTER 2: Reorienting Scientific Explanation – The Case from Robustness in Biological Systems

2.1 Introduction

To paraphrase the ecologist C. S. Holling: There are two ways to look at the world. In one way of looking at the world we see how changes cascade through an otherwise stable world. Winding a watch causes the clocks hands to move, death follows poisoning, and a sprout grows after watering a seed. But another view of the world concentrates not so much on identifying important changes or the presence or absence of an effect, as upon the persistence of things and the relations that maintain them. In this way of looking at the world we see how stability and function are maintained in a world that is constantly in flux. An ecosystem maintains the same yearly cycles of community structure across seasonal variation, species retain their general character as the individuals that make them up are born and die, a cell maintains homeostasis across a wide range of environments.

These are two different ways of viewing the behavior of processes and systems, and the usefulness of either view depends greatly on the properties of the system concerned. If we are examining a device that performs a specific function in a reliable and nonvariable way, then we are likely to be concerned with identifying the factors that make a difference to whether or not that device performs its function. But if we are examining a system that is profoundly affected by a fluctuating environment, such that the way it operates varies in response to unpredictable external changes, then identifying the factors that change its behavior become less important and we are likely to focus more on the persistence of the internal relations responsible for its function. Attention shifts from a focus on the causes of change to explaining stability in response to constant change.

One striking feature of many biological processes and systems is their high degree of robustness, or the reliable production of a similar output under a wide range of circumstances. Despite the ubiquity and importance of robustness, many explanations pass over the organizational details that give rise to it. Biological systems are often self-organizing systems full of complex nested relations and interactions that resist decomposition into discrete parts that can be modeled in a straightforward way. Different and useful insight might be obtained, therefore, by focusing on the relations that generate robustness at the system level, and by shifting emphasis from change-dependent relations to the holistic relations that maintain the function or behavior of a system.

The overall goal of this paper is to explore the behavior and organization of robust systems in order to see if a different perspective can yield insights useful for both theory and practice. I start by discussing two fields that have recently started to pay increased attention to robustness and the mechanisms that generate it: ecology and evolutionary developmental genetics. It is worth going through developments in both fields because an increased attention on robustness occurred independently and probably for different reasons. I then turn to the concept of robustness directly. Merely noting whether or not a system is robust is often explanatorily inadequate, so I propose that our focus should shift to the different ways in which a system can be robust. I develop a general taxonomy of kinds of robustness in order to help with this goal. I conclude by returning to the idea that there are two complementary explanatory perspectives that are useful for understanding the world.

It is worth noting before we delve in that the terminology surrounding the concept of robustness is a mess, likely an effect of the rapid and recent increase of interest in the topic. Terminology is inconsistent within fields and difficult to compare between fields. I will do my

best to simplify and clarify the terminology in this section, and I will later define terms for how I will use them when developing a more generalized framework for understanding the varieties of robustness.

2.2 Two Explanatory Reorientations in Biology

2.2.1 Dynamics and Resiliency in Ecology

Ecology is the study of the interactions between organisms and their environment. Early theoretical approaches to ecology were quantitative in nature and focused on the dynamics of organismal relations. Vito Volterra, a mathematician and physicist, was an important figure here. He helped develop a mathematical approach to the study of biology, his most famous contribution being the Lotka-Volterra, or predator-prey, equations. These are a pair of differential equations used to describe the dynamics of a two species interaction where one is a predator and the other is prey. Quantitative approaches were soon developed for other ecological interactions and processes: birth and death rates, immigration and emigration, survival and reproduction, population growth rates, density-dependent interactions, energy/food intake, competition, disease, parasitism, mutualism, etc. A common feature of this approach is that it isolates a relation of interest and treats it like a dynamical system that can be described through mathematics. This is a powerful approach, because if the system is modeled accurately the model gives insight into the causal relations that are important for explaining the phenomenon of interest. It will also show the effects of changing those relations, of tweaking the variables, so to speak.

Knowledge of ecological relations is critically important because human well-being depends on a wide range of ecosystem services. For example, fisheries provide a significant

proportion of our food. Understanding the dynamics of fish populations allows us to figure out how many fish we can catch, when we should catch them, and at what size while still maintaining the population at a level that will allow us to keep on fishing indefinitely. It turns out though, that the history of ecosystem management is primarily a history of failures to predict the effects of change. Whether that be the effects of harvesting, the introduction or removal of a species, a natural disaster, or the spatial fragmentation of a population.

Given a century of advances in dynamical theory, why are we still not very good at predicting the effects of change on ecosystems? A central problem is that ecosystems are complex. They are often characterized by multiple possible outcomes and the potential for rapid change and major regime shifts due to slower and smaller changes in internal or external influences. Changes in interactions at small scales can eventually end up shaping the dynamics of the macroscopic system which then can feed back to reshape the interactions at the smaller scales. Most ecosystems are also likely to be in a transient state, which makes them behave quite differently than a system near equilibrium modeled by dynamic equations.

Another problem, emphasized in Connell and Ghedini (2015), is that there is a predisposition to the way we study change that comes from the way we think about disturbance. A disturbance is considered to have occurred when some event causes observable changes in some property of interest; for example, species composition, biomass, or productivity. Conversely, if we notice that there is no observable change in the property of interest, we conclude that the disturbance had no effect, and thus wasn't a disturbance at all. This approach ignores any reorganization or compensatory processes that might be active within the system that buffer against change. The problem of prediction and insights into the non-equilibrium and complex character of most ecosystems led some ecologists to push for a different perspective for

understanding ecosystem behavior.

This alternative approach, which goes by many names but is probably most often called resilience theory, was introduced into the ecological literature by Holling (1973). According to this theory change is ubiquitous in ecosystems. In addition to analyzing the component relations of an ecosystem and combining them to try and build a bottom-up understanding of system behavior, we should take a top-down approach and analyze the properties that determine the persistence of relationships within a system, as well as its capacity to absorb change and maintain structure and functions. Central features of the resilience approach are the emphases on alternative stable states for a system and the mechanisms that are responsible for renewal and reorganization before, during and after a disturbance.

The resilience framework became especially important in two projects. It was used in ecosystem management and conservation because it helped to predict ecosystem recovery after a disturbance which in turn is used to shape policy for resource harvesting and responding to environmental disasters (Levin and Lubchenko 2008). Elsewhere in ecology, there was a drive to discover the mechanisms and relationships that underpin ecosystem resilience in order to predict responses to future internal or external changes.

Resilience theory has undergone a lot of refinement in the last four decades as insight has been drawn from many fields. An unfortunate side effect of this interdisciplinary effort has been an explosion of conflicting terminology. There is considerable confusion about the meanings of terms associated with resilience theory.

For simplicity, I will draw on Hodgson et al. (2015) while providing a more detailed overview of the approach. The terminology and conceptual breakdown presented here is by no

means representative of the entire field, but it is clear and straightforward. The resilience of an ecosystem is a measure of the capacity of that system to persist or maintain function in the face of disturbance. This capacity can be broken into two components, resistance and recovery, which can vary (possibly independently) across systems. Resistance describes the immediate impact of an external disturbance on an ecosystem, while recovery represents the capacity of a disturbed ecosystem to return towards an equilibrium. The rate of return is called elasticity. If alternative stable states exist for the system, then latitude describes the distance between a stable state and the tipping point into a different stable state. Different systems can vary across all these parameters.

A researcher might measure the level of disease exposure that is required to cause infection (resistance), the capacity to fight off an infection and return to a healthy state (recovery), the rate of recovery (elasticity), and the severity of disease that is required to push the system into another permanent state like death (latitude). As should be obvious, resilience is not a universal character of a system. A particular system might be more resilient to one type of disturbance and more susceptible to another, just like a particular person might be resilient to one disease and susceptible to another. Furthermore, the resilience of a system to a particular disturbance can change as the system evolves, often in subtle ways. The idea that the impacts of disturbance might depend on initial system state is known as hysteresis. Hysteresis is well known in ecological systems. This implies a very interesting feedback loop between disturbances and resilience, mediated by the adaptive dynamics of the system (Hodgeson et al. 2016).

The approach taken from physics and engineering of breaking a system into parts and describing them is insufficient for understanding the robustness of complex adaptive systems like ecosystems, but it can provide insights into what makes ecosystems robust and how. Indeed, a

fundamental challenge for understanding and managing complex systems is to unite the top-down system level approach with the bottom-up mechanism approach. (Levin and Lubchenko 2008). This two-prong approach is important for increasing our predictive power when moving towards novel future changes like climate change.

2.2.2 Mechanisms and Robustness in the Cell

Evolutionary developmental biology (evo-devo) is the study of how developmental processes evolve with a focus on the dynamics of gene expression, the complex relations between genotypes and phenotypes, and how everything together affects phenotypic evolution. Evo-devo only came together as a field in the late 20th century after advances in molecular biology and the widespread sequencing of DNA, RNA and proteins led to a better understanding of the molecular basis of developmental mechanisms.

The questions being addressed by evo-devo have a longer history. Population genetics and quantitative genetics produced a large body of knowledge regarding the evolution of genotypes and phenotypes but had to make very simple assumptions about the relations between genotypes and phenotypes. Scientists for the most part had to rely on statistical correlations among quantitative characters, i.e. those traits that varied in such a way that they could be measured. Understanding the evolution of complex genotype-phenotype relations remained elusive. Evo-devo tackles these complex relations head-on.

Investigations into the developmental pathways underlying phenotypes have revealed a high degree of complexity in developmental systems. Even some of the simplest organismal phenotypes are influenced by hundreds of genes. Developmental pathways are often also exceedingly robust (Nijhout 2002; Bateson and Gluckman 2012). In the broad domain of

organismal biology, robustness usually refers to the ability of a system feature to maintain its function against perturbations. System features of relevance include: molecular structure, an enzyme's catalytic ability, a regulatory circuit's expression, a metabolic network's activity, or the expression of a phenotype at the organismal level. Perturbations are often categorized as coming from three sources: stochastic noise, environmental variation, and genetic variation[11].

A flurry of research into robustness in recent years has found that proteins can tolerate thousands of amino acid changes, metabolic networks can function after the removal of important chemical reactions, gene regulation networks can withstand changes to key gene interactions, and genes can be deleted without affecting an organism's adult phenotype (See Wagner 2005; Steinacher and Soyer 2012; Whitacre 2012 for more details). This massive amount of observed robustness has spurred investigation into the mechanisms underlying and responsible for robustness, the evolution of robustness and the evolutionary reasons for its ubiquity. This has also led to a great deal of interdisciplinary crossover with systems biology.

What is the relevance of this outgrowth of research to the project at hand? Because of the complexity of developmental systems, correlations between singular changes or interventions on a system and downstream effects give very little information about the causal relations connecting the two. Strong correlations might be primarily an experimental artifact if causal connections only hold in very constrained circumstances as the result of being the product of

[11] Noise refers to the stochastic fluctuations that occur in any biological system, for example in the concentration of a biological molecule or in a cell's position, either over time, or between two genetically identical individuals, even if the external environment is constant. The second kind of perturbation is variation in the external environment, for example a change in temperature, salinity or nutrient availability. The third kind of perturbation is genetic change, either through mutation or recombination. Felix and Wagner (2008).

complex and rare interactions. Conversely, there might be no correlation between a particular change and the system's response, not because the two are not causally connected, but because of compensatory effects resulting from the robust nature of the causal connections. Worries about recovering causal structures from correlative data are not unique to biological systems, but they do cause particular trouble here.

The practical application of a bottom-up approach that investigates a system mechanism-by-mechanism and relation-by-relation to the investigation of entire developmental and cellular pathways has been limited. This is mainly due to the enormous degree of dimensionality and non-linearity found in biological systems. Intervening on a particular mechanism to measure perturbations in the system for all these dimensions and their interactions is impractical. Constraining the complexity of an explanation by simplifying the number of factors in play through the judicial use of ceteris paribus clauses might lead a false picture of the causal complexity of the system. If system decomposition and bottom-up causal-mechanical explanations are too complex and bloated, or not widely generalizable, what can be done? Like we saw in ecology, the answer has been to investigate and characterize system-level properties in addition to the lower level mechanisms.

While a full theory of robustness in cellular systems remains elusive; both in terms of the mechanisms by which robustness is achieved and the selection pressures that have led to widespread robustness, there has been progress in uncovering systems principles that contribute to robustness across biological disciplines[12]. Important system-level properties include

[12] Whitacre (2012) and Bateson and Gluckman (2012) provide good overviews of some of this progress.

redundancy and various robust network architectures--which Wagner (2005) calls distributed robustness. Redundancy refers to two parts of a system being able to perform similar or identical tasks. This might be because the parts are identical, such as in the case of kidneys or as the result of gene duplication events. It might also be because two different parts can perform the same task. This is usually called functional redundancy. Think of a hybrid car with its two power sources. In the cell, glycolysis and lactic acid fermentation are functionally redundant in regards to the generation of ATP.

Robust network architectures include scale-free networks, small world networks and bow-tie architectures (Kitano 2004, 2007). The details of these network structures are not important for now. They can be described in a general way as when the parts of a system contribute to system function, but all of these parts have different roles. When one part fails or is changed through mutations, the system can compensate for this failure, but not because a ‘‘back-up’’ redundant part takes over the failed part’s role. Wagner (2005) calls this robustness that is due to the complex network of relations in a system ‘distributed robustness’, and gives another helpful example: “On lower levels of biological organization, the paradigmatic examples of distributed robustness are biological macromolecules, such as proteins and RNA. Their folding in space involves the cooperation of hundreds of building blocks—nucleotides or amino acids—each of which forms unique (non-redundant) interactions with other building blocks. If a change in one amino acid leaves a protein’s three-dimensional conformation unchanged, is this because the protein contained a redundant ‘back-up’ amino acid? Clearly not: The thermodynamic stability of any macromolecular structure is the sum of all contributions that such interactions make to the structure’s free energy.” (Wagner 2005).

Research into the causes, mechanisms, and organizational features of robustness is

changing our understanding of ecology, developmental, and evolutionary biology. Whitacre (2012) speculates that robustness may be ubiquitous in biological systems in part because proteins, cells, biochemical networks, immune systems, organisms, and natural populations exist within changing and sometimes novel conditions under which the maintenance of satisfactory performance will determine persistence or function. No matter the reason why robustness is so commonly found in biological systems at all levels, it is clear that it is an important feature that calls out to be understood. It is also changing the way that biologists approach their research and the explanations they provide. As we have seen in these examples, researchers have begun to focus on the complex interrelations that exist at the system level in order to supplement the investigations into mechanistic approaches used for understanding the individual parts of a system.

2.3 Robustness in Philosophy

Discussions of robustness are included in most causal explanatory accounts, but it is a topic rarely treated at length. There has been a relative consensus on the informative nature of robustness. The stock example regards the supposed inevitability of the First World War:

> The stability of an outcome of a causal process in spite of significant variation in initial conditions can be informative about an ideal causal explanatory text in the same way that it is informative to learn, regarding a given causal explanation of the First World War, that a world war would have come about (according to this explanation) even if no bomb had exploded in Sarajevo. This sort of robustness or resilience of a process is important to grasp in coming to know explanations based upon it. (Railton 1981: 251)

On some accounts, an explanation showing that a process is robust is of increased

practical value, rather than increased explanatory value. Railton (1981), Wimsatt (1981), and Batterman (2002) roughly fall into this tradition. Others, such as Jackson and Pettit (1992), Sterelny (1996), and Strevens (2008) see the recognition of robustness as increasing explanatory value. All of these accounts share a common feature. They are only concerned with whether a causal process or system *is robust at all* relative to some change or difference.

I will instead offer a framework for distinguishing the importantly different *ways in which a causal process or system can be robust*. Woodward (2013) makes some first moves in this direction when he discusses biological robustness as an example when arguing against the overextension of the mechanistic explanation program. There distinguishes between robustness stemming from redundant back-up mechanisms, robustness stemming from complex dynamics (what he calls robustness without fine-tunedness), and robustness resulting from topological or network features of the system. I will not argue here for the necessity of a fuller account of robustness in a general theory of scientific explanation. Instead I will just contend that if a theory of explanation cannot make room for a focus on system-level details like those found in my analysis of robustness, then it cannot be adequate for understanding causal reasoning in biology.

I pause to note the relations between my project and two similar projects by scientists, e.g., Bateson and Gluckman (2012) and Whitacre (2012). Bateson and Gluckman review the interaction between plasticity and robustness, which they define as the "consistency of the phenotype despite environmental or genetic perturbation" (Bateson and Gluckman 2012, pg. 8). They use a more delimited notion of robustness, constrained as it is to phenotype consistency across developmental environments, than the one I will try to develop. There is a similarity though in that they are also particularly interested in the many mechanisms that are responsible for generating robustness. They survey, via a list of definitions, some of the concepts that are

important for understanding how different kinds of robustness are produced in the context of the evolution and functioning of developmental and behavioral systems. Whitacre (2012) also recognizes that there is an assortment of mechanisms that underlie robustness, but highlights how many of these mechanisms share similar system-level properties, such as modularity, small-world networks, bow-tie architectures, and other topological features. He then explores the relation between these system-level features and common features of the processes that generate them, such as functional redundancy, degeneracy, adaptive plasticity, environment tracking, and environment shaping. Neither paper goes as far as trying to develop a more general hierarchy of broad sub-types of robustness, which is what I lay out in the next two sections. I don't see my project as in conflict with either Bateson and Gluckman (2012) or Whitacre (2012). They provide some of the necessary details for understanding robustness in biology. I am trying to subsume those details into a more general taxonomy of robustness in order to suggest that paying attention to how and why a system is robust is a fruitful perspective for understanding the behavior of systems in domains other than ecology, genetics, and cellular biology.

2.4 Ways to be Robust

I will start my account of robustness with the related notion of reliability. A *reliable* process is a process that consistently produces some phenomenon (or suitably similar small set of phenomena) starting from some set of initial conditions (or suitably similar small range of sets of initial conditions). It is essentially a regularity relation linking two phenomena. Whether we take a process to be reliable or not will depend on the relation between the process and the target of explanation. The process of flipping a fair coin isn't a reliable process for generating heads, but it is a reliable process for generating a 50/50 distribution of heads and tails over an extended series of flips. A reliable process has the general form of [Initial Conditions] --- [Black Box Process] --

--> [Regular Phenomena]. A thorough explanation will try to open up and show what is inside that black box.

Robustness is a different type of relation. A relation is robust if it holds under a range of circumstances. This might be a relation between two different events or entities, or it might be a relation across the same entity at different times. More specifically, a process is *robust* to the degree that it can reliably produce a particular outcome under a wide range of circumstances. Namely, a wide range of initial conditions, perturbations, and parameters (cf. Strevens 2008). I will also be speaking of the robustness of systems. A system is robust to the degree that it can maintain its function or behavior under a wide range of external or internal perturbations (cf. Kitano 2004; Woodward 2013). This is just a way of rearranging the above criteria, in both cases the key feature is stability in the face of changes. I won't make sharp distinctions here between processes and systems. I view systems in this context as a set of integrated causal processes interacting in a structured way to maintain an equilibrium point or to produce a behavior or product. A mechanism is an example of a type of system structure that produces something. Ecosystems and organisms are systems that maintain equilibrium points.

A robust process/system is a reliable process/system, but a reliable process/system is not necessarily robust. For example, a retinal scanner reliably unlocks a door given the right eye input, but it is not robust to variation in inputs in regards to door unlocking. Nor would the owner of the device want it to be. A suitably advanced one might be so sensitive to initial conditions as to lock-down if the temperature, pressure, gravity, etc. varies at all. Yet, given the right initial conditions, the door unlocks every time. Conversely, the same system might be extremely robust in regards to the behavior of keeping a door locked.

2.5 Varieties of Robustness

I distinguish among three different ways in which robustness can be achieved and use the terms *redundancy, resistance, and resilience* for the respective varieties of robustness. As we have already seen these terms are used in different ways across a wide variety of fields, so it will be necessary to stipulate their meanings. The way that I will use these terms will cut across some of the literature, both scientific and philosophical. My primary goal in this section is not to try and fix the meaning of these terms. Rather, I am attempting to get at a cluster of related and often confused concepts in order to articulate them in a clear way. Figure 1 at the end of this section gives a pictorial version of the terms I use in my account.[13]

Before looking at them each in more detail, it might be easiest to quickly compare and contrast the three kinds of robustness. Here is a rough analogy to get us started. We have just been through a bad storm and most of the trees at the park have been destroyed. Three types of trees survive: a large oak tree, a willow tree, and an aspen tree from a clonal colony. We want to know what explains the tree survival in regards to the storm. How where they able to remain intact? Luckily, there was a video camera trained on the park. As we watch the tape we see the wind slowly ramp up. In response to the wind, the large oak tree remains relatively unaffected. It maintains its form in the face of the howling winds. The flexible willow tree, on the other hand, whips about and bends in the wind. It changes its form in response to the storm in order to stay intact. The aspen clones form a small colony cluster connected by a common root system. The individual aspen clones are not as strong as the oak nor as flexible as the willow, and one by one

[13] Whitacre (2012) was an influence on my thinking here and the paper provides a good overview of specific examples of some mechanisms underlying robustness in biological phenomena, with a focus on developmental and cellular biology.

they start to snap. But, one single tree makes it through the storm and so the colony survives.

In this analogy, all three tree varieties are robust to large changes in wind speed. Extending the analogy then we see three different ways to be robust to the storm. When the response to change is *resistant*, it is because either the system or process itself isn't affected, or if it is affected it doesn't change the character of how the process unfolds. The oak tree is insensitive to the variability in conditions and so is resistant. It looks exactly the same under a wide range of wind speeds. This contrasts with *resilience*, where the system is changed or reorganized but still maintains its function or ability to bounce back to equilibrium after the disturbance. The willow tree responds dynamically to the variability in conditions by flexing and changing in response to the wind and so is resilient. It looks very different in different wind patterns. Finally, robustness takes the form of *redundancy* when there are back-up processes or fail-safes that can kick in if one part of the system fails. The aspen colony survives because even though individual aspen clones are destroyed by the storm, there are duplicate back-ups that make it. A pictorial overview of these concepts can be found below.

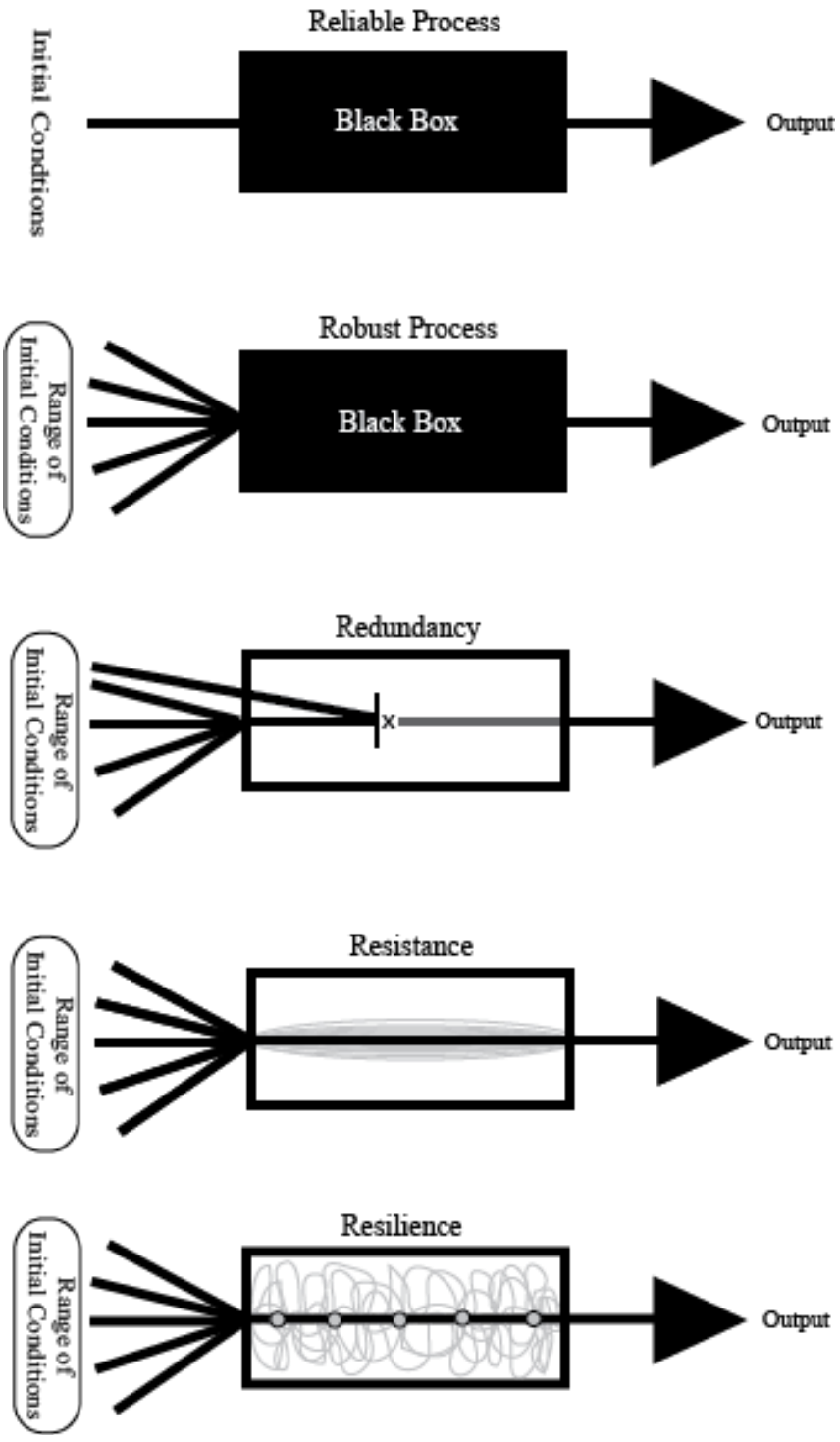

Figure 2.1 *Illustration of robustness categories*. In the redundancy illustration, the process is disrupted and a back-up takes over. In the resistance and resilience diagrams, the black line represents the mean response to disturbance or changes in initial conditions. Grey lines represent individual paths through state space in response to particular changes. Grey dots represent basins of attraction or selection events, if they exist. Note that resistant processes don't change their behavior to a high degree, while resilient systems are marked by a high degree of flexibility.

Robustness isn't a universal property of a given system or process. It is always relative to both a particular function or behavior, and a particular kind of disturbance. Systems might also be robust in different ways to different things, such that they are resistant to some changes and resilient to others. Similarly, robustness is often nested. A resilient process might be made up of a network of sub-processes that are each themselves resistant, or have subsystems full of redundant parts. With these rough distinctions in mind, I will now look at each kind of robustness in more detail.

2.5.1 Redundancy

Robustness is achieved through redundancy when a system or process has multiple means by which to produce an effect or behavior. If one process or structure is disrupted, another process or structure ensures that the whole system isn't disrupted. I distinguish between two general kinds of redundancy: backups and compensatory processes.

Back-up processes enhance robustness because they can work as fail-safes. If one process fails, then it can be replaced by others. Back-ups can come in the form of duplication, when there are, strictly speaking, more of something than are necessary for the task. Think of kidneys, or the load-bearing beams of a bridge. Back-ups can also come from functional redundancy in a diverse set of processes where each process produces the effect in a different way. Take an effect *e* and initial condition sets *a*, *b*, and *c*. A mechanism composed of three distinct causal pathways for bringing about *e*--call them A, B, and C--where each pathway is sensitive to a different set of initial conditions, *a*, *b*, and *c* respectively, would be functionally redundant. If one of the processes is disrupted, the others can still bring about *e*. Think of a car that has both seatbelts and airbags, or the different cellular pathways for transforming glucose into usable energy under different conditions.

Finally, take a related case where there is a system containing processes A and B. A has two effects, e_1 and e_2, and B has two effects, e_2 and e_3. These are distinct processes that do different things, but relative to e_2, the processes are functionally redundant. A is a partial back-up for B, and vice versa. This is often called degeneracy. Many biological processes that are functionally redundant are also degenerate.

Redundancy can also come in the form of compensatory processes. Instead of acting as an independent back-up, one process responds to the disruption of a different process in order to bring about the effect. Repair mechanisms are a good example of a compensatory process. Heat shock protein Hsp90 is another important process. Hsp90 is a protein that confers robustness to the process of protein folding by assisting other proteins to fold or refold into functionally relevant conformations when temperatures are elevated above physiological conditions.

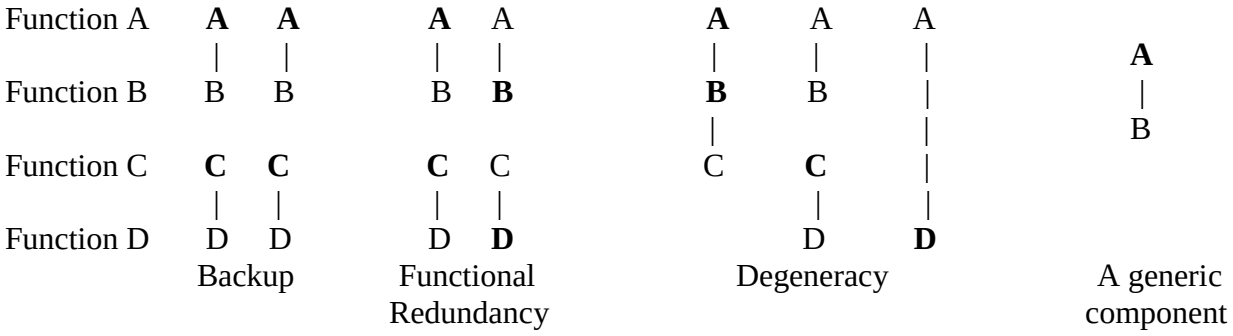

Figure 2.2 *Illustration of Types of Redundancy* - Component parts of a system often display a range of closely related functions. Component parts are represented in the figure as linked letters, with the letters representing different functions. Bolding indicates **primary** function. When two or more components and their functions are identical, they form redundant backups of each other. Components that differ in structure, but not function, display functional redundancy. A series of components that back-up some, but not all, of each other's functions display degeneracy.

2.5.2 Resistance

Robust processes that are *resistant* are insensitive to change. Robustness is achieved

through resistance when *the way in which a causal process unfolds* is largely unaffected by changes in initial conditions or background conditions. A system is resistant when it maintains stability near an equilibrium steady state because it is insensitive to perturbations to the system.

Of all of the types of robust processes and systems that we encounter in our day to day life, we might be most conceptually familiar with the resistant ones. Stable rigid structures like steel beams, brick houses, and mechanical watches composed of intricate gears are all resistant insofar as they are robust to external disturbances. When a response to an external change is resistant, it is because either the process itself isn't affected, or if it is affected it doesn't change the character of how the process unfolds. This contrasts with redundancy, where a back-up process kicks in or takes over, and resilience, where the system is changed but can still maintain its function or bounce back to equilibrium. Microscopic organisms called tardigrades, or water bears, are a fascinating example of biological resistance at the organism level. When they are in a suspended metabolic state they can persist for years under tremendous extremes in temperature (-200 to 150 C), pressure (vacuums to 1200 atm), dehydration, and radiation (Whitacre 2012).

Resistance is a characteristic of systems that are robust to particular changes because those changes don't affect the character or function of the system. One reason is because the strength of the property or force that is responsible for the stability of the system dominates changes in the strength of the external forces that typically interact with it. Think again of a steel beam. Steel beams are largely unaffected by changes to their environment. The chemical bonds that maintain steel are strong enough that they are not broken by the changes in temperature, gravity, kinetic forces, etc. that steel normally encounters. This same principle applies to the oak tree in my previous example. Related to this notion of inactive resistance to strength, is resistance due to a system's buffering capacity. Rather than being unaffected by disturbances up

to a certain magnitude, a system might be able to absorb a certain amount of disturbance without being affected. Systems that are robust only because of a single dominating force or property also tend to fail catastrophically if that force is overcome because there is nothing else to maintain the system's structure.

Finally, structural features might limit accessibility to the more vulnerable aspects of a system. Hard shells in many biological organisms help them resist physical forces and dehydration. The opaque casing of a lock helps make it resistant to picking because it hides information about how the lock works. Limited and hard to predict emergence times in cicada insect populations make them resistant to extinction.

2.5.3 Resilience

Resilient processes and systems are characterized by their flexibility. Robustness is achieved through resilience when there is reliability, but *the way in which a causal process unfold*s varies considerably in response to differences or changes in initial or background conditions. Resilient processes are often marked by their punctuated movement between basins of attraction in response to disturbances. A resilient system is robust because it is able to maintain functionality or overall behavior by (1) reorganizing or switching between alternative stable states or behavioral regimes, or (2) by harnessing the incoming stochastic variation itself to produce a behavior or maintain functionality.

Complex systems that are resilient have many parts that contribute to system function, but all of these parts have different roles, and may only weakly interact with each other. When one part fails or is taken out, the system can compensate for this failure, not by a redundant part taking over for the failed part, but by reorganizing or compensating in other ways. For example,

there are cellular metabolic networks containing hundreds of products and reactions with no redundancy of reaction pathways that are still extremely robust to the elimination of (normally) central chemical reactions. When individual reactions are eliminated, the networks respond by spontaneously reorganizing and begin to shuttle the flow of matter through parts of the metabolic network that were not eliminated.[14]

Wagner (2005) discusses another example dealing with gene regulation networks. A study of a fly development model focused on a phenotype that is the product of gene network containing at least 48 non-redundant biochemical parameters. The researchers found that these parameters could be changed by an order of magnitude and the network would still maintain its function. The source of the robustness was again due to the distributed architecture of the network. In this case, and the one above, the specific path through the network that is taken depends on the initial conditions and how the network has been disrupted. But no matter the path taken, the output is the same. Furthermore, the large numbers of distinct, contextually active, regulatory interactions make it impossible to reduce the origins of this robustness into simple sets of independent processes, despite the rather simple individual mechanisms making up the network.[15]

A second variety of resilience is the result of the stochastic nature of much of the biological world. Many biological processes are not deterministic at the level of organization at which they operate. By this I mean that there is not a one-to-one matching between specific input conditions and the way the process unfolds in order to produce the output of the system. Instead

---

[14] See Wagner (2005) for more details on the biology.
[15] A point emphasized by Whitacre (2012).

the process proceeds and changes in a stochastic fashion. This is different from the example discussed above. There, differences in the input change how the process unfolds, but this doesn't preclude that there is a one-to-one matching between specific inputs and how the process unfolds. There is process variation between inputs, but not necessarily within replicates of the same input.

The behavior of processes that take stochastic variation and reliably transform it into robust end states come in at least two interesting types. The first is exploratory behavior, which involves repeated selection over a series of subsystem states that are each elaborations of previously selected states. Exploratory behavior thus requires one process that generates variety and another process that selects from that variety so it can be fed back in and expanded upon. Selection may be positive (activation or reinforcement) or negative (inhibition or culling). Evolution by natural selection is an open-ended process of this type. There are a variety of biological systems that exhibit exploratory behavior and have a selection mechanism that molds the behavior for a particular function. When this is the case we can get the robust relation between input and output even though the process that links the two is not deterministic and may never look exactly the same. Examples of resilient systems that include exploratory behavior include: immune systems; microtubule spindle formation; cellular differentiation, proliferation and apoptosis during development; chemotaxis and cell motility; the formation of neural connections and vascular growth during development; and behavioral adaptation in complex species such as predator avoidance, habitat tracking, and adaptive foraging.[16]

The second type of behavior that harnesses stochastic variation is characterized by

[16] Some examples taken from Whitacre (2012).

random walks and basins of attraction. This type of behavior involves a system that changes over time according to a random walk. As the system explores the state space open to it, it will occasionally happen upon important functional states or structural conformations, or be pulled into a local basin of attraction that will cause the system to drift towards the attractor over time. By randomly moving through a series of functional states or structural changes that change the state space open to the system in a directed way (and thus order the functional states) the system can be made to robustly produce an effect or behavior given enough time. Similarly, a system can robustly produce an effect or behavior if the basins of attraction that the system moves through tend to be ordered in some way. Many cellular processes that are subject to thermodynamic noise (Brownian motion) behave in this manner[17]. Some robust ecosystems that are able to return to equilibrium after a disruption might also be fruitfully described in this way. In this description, the system is knocked away from equilibrium via stochastic disruptions, and the equilibrium point acts a single global attractor that brings the system back to it over time.

Reviewing, I have tried to show that there are three distinct varieties of robustness: redundancy, resistance, and resilience. These categories cut across other discussions of robustness in the scientific and philosophical literature to various degrees. Ecology tends to focus on resilience and resistance, with some discussion of redundancy in the mechanisms that lead to resilience. Evo-devo tends to focus on varieties of what I call redundancy and resilience, though it almost never uses the term resilience. The philosophical literature on robustness is almost completely silent on any subdivisions of the concept.

[17] See Skillings (2015) for a more detailed look at stochastic mechanisms of this type, such as ribosome synthesis.

Robustness isn't a universal property of a given system or process. When attributing robustness to a system or process, it is always relative to a particular explanandum and to a particular variety of disturbance. Systems might also be robust in different ways such that they are resistant to some changes, resilient to others. Robustness can also be nested in that a system can be robust in different ways at different levels. A resilient process might be made up of a network of sub-processes that are each themselves resistant, or have subsystems full of redundant parts.

Using the notion of nested systems or processes it is possible to describe the differences between redundancy, resistance, and resilience in another way. Breaking a system down into subsystems, or a process into sub-processes, and comparing the reliability and relations between the components to the reliability of the whole will help determine the nature of the robustness of the overall system. For example, in a resistant system the reliability of the entire system will be at best as reliable as the least reliable part. This is because the system will break down, if any of the parts break down. The chain is only as strong as its weakest link.

2.6 Reorienting Towards Robustness

The beginnings of modern science emerged out of the successful and fruitful marriage between physics and mathematics in the sixteenth and seventeenth centuries. The traditions of analysis and explanation that later developed in the special sciences were largely inherited from physics and its applied variants. A tendency to quantify and emphasize the machine-like nature of the world was to a large degree inherited along with it. It is important in classical physics and its precursors to redescribe a phenomena in terms of a set of properties and then describe the invariant relations between the properties that guide the behavior of the whole system. It is similarly important to identify the dominating forces underlying the production of some

phenomena, while ignoring those factors that made little to no difference. This type of analysis lends itself to a picture of the world where everything is pushed and pulled around in a quite orderly and discrete manner. But this orientation may simply be a remnant of an investigative approach developed in one area because it was productive and then transferred to another where it may not always be as useful. Our traditional view of natural systems, therefore, might well be less a meaningful reality than a perceptual convenience.

Focusing on features like robustness is one way to cash out the perspective that emphasizes the relations that maintain stability in a world of flux. Looking closely at the details of scientific practice we find compelling support for highlighting the organizational details that give rise to robustness. Shifting to this perspective is a powerful tool for gaining a greater understanding of the world. Moving beyond these observations, I have suggested a framework for thinking about robustness. Applications of this framework and perspective may also be helpful with understanding and thinking through other puzzling cases.

The approach inherited from physics of breaking a system into parts and describing them is insufficient for understanding complex adaptive systems like those found in biology, but it can provide insights into what makes these systems robust and how. Indeed, a fundamental challenge for understanding and managing complex systems is to unite the top-down system-level approach with the more standard bottom-up mechanistic approach. This two-prong approach is important for increasing our predictive power when moving towards novel future changes such as ecosystem response to climate change, or using knowledge of developmental processes in one group to understand how they work elsewhere.

CHAPTER 3: From an Entangled Bank – Translating and Interpreting Causation in Biology

3.1 Introduction

The universe is a vast weave of causal influence. For any given event there are a bafflingly large number of causal interactions involved in bringing about that event. One goal in the philosophy of causation is to parse this vast network of interactions so that we can pick out tractable causal descriptions and generate useful causal explanations[18]. Useful in the sense that they increase our understanding of broader underlying patterns and serve to increase our predictive powers or control of the world. Many biological systems differ in important ways from the physical and chemical systems used as examples in most accounts of explanatory practice or causal reasoning. Understanding causation in biology comes with a new set of challenges due largely in part to the fact that the biological world is not only complex, but highly organized, full of recursion, often historically contingent, and without many widely generalizable features. On some accounts, biological systems realize a new emergent and distinctive causal regime beyond those found in purely physical systems (Mossio et al. 2013; Mossio and Bich 2014). In any case, biologists have had to expand their toolset for causal reasoning in order to deal with peculiarities in the behavior of biological systems.

In this chapter I aim to show how approaches to causal reasoning that have been modeled on the investigatory procedures that have been successful in the physical sciences (introduced in

[18] My project here takes a view of causal reasoning in the spirit of what Woodward (2014) calls a functional approach to causation. As opposed to projects in the philosophy of causation that are focused on metaphysics, descriptions of ordinary usage, or how causal claims fit with fundamental physics, the functional approach evaluates causation in terms of its usefulness for particular projects, such as explanation or description.

Section 3.2) can, in some situations, obscure the complex causal relations present in many biological systems. The situation can be especially problematic when exporting causal explanations for use in different domains or at other scales. I introduce two related problems in this paper: the *causal translation problem* and the *causal interpretation problem*. I believe these problems are best illustrated by working through biological examples in detail.

I introduce the *causal translation problem* in Section 3.3 by contrasting the causal features of the "nanoworld" inside living cells with the causal features of our everyday world--the world of "moderate-sized specimens of dry goods", as J.L. Austin puts it (Austin 1962: 8). The *translation problem* refers to the mismatch that comes from applying causal reasoning strategies that are successful in one domain to a different inappropriate domain. The resulting problems stem from the fact that the causal interactions in the systems of interest differ too much and so make comparisons or analogies between them only marginally useful, if not downright misleading.

The *interpretation problem* has to do with the difficulty in applying the simplifying assumptions necessary for tractability while still adequately capturing the causal structures of complex systems (Section 3.4). The tradeoff between tractability and adequacy can make it difficult to interpret the underlying causal structure of the system of interest. Problems emerge when oversimplified accounts are treated as accurate depictions, more often than not by non-specialists using those accounts for other purposes. I illustrate this problem by examining work on complex biological systems featuring properties such as low modularity, recursive hierarchical organization, self-maintenance, self-production, and high degrees of robustness.

3.2 History and Approaches

The contours of the history of debates about causation and scientific explanation are well explored in the philosophical and historical literature (references). In this chapter I will focus on some of the conceptual tools useful for causal reasoning. Causal reasoning about the world depends on finding a way to pick out the most salient cause(s) of the phenomenon of interest. This process takes many forms, including but not limited to: abstraction (the removal of irrelevant details), idealization (the introduction of explanatorily beneficial falsehoods), reduction (focusing on smaller or more fundamental parts), and decomposition (breaking down a system into modular constituent parts). Simplification via these methods is useful because they help with tractability, can increase our control of the world by revealing ways to effectively intervene on it, can aid in prediction, can reveal underlying generalities, and can help with understanding.

Causal reasoning in the special sciences emerged out of a scientific tradition of analysis and explanation that held up mathematical and dynamical models of phenomena as exemplars[19]. Mathematical approaches have been the most successful in the physical sciences, especially mechanics and chemistry. This is in part due to the fact that there are a lot of phenomena in these fields that allow for strong simplifying assumptions. This may be due to the fact that the behavior of the system of interest is dominated by relatively few forces or properties (e.g., inertia, friction, density, etc.) or that the system is homogenous and/or is analyzed at the scale where the micro-behavior of its components can be averaged out (e.g., pressure, temperature,

[19] In this chapter I don't make a strong distinction between causal reasoning and causal explanation. I take it that a particular form of causal reasoning is justified insofar as it can be used produce adequate causal explanations.

speed, stoichiometry, etc.). Systems with such properties can be modeled in highly abstract or idealized ways, often resulting in models that are easily understood yet still offer a great deal of predictive power and an increased control over the world.[20]

As the complexity of the system increases, this (relatively) straightforward approach to science begins to be less practical and less successful.[21] Relative to physics and chemistry, biological systems are made up of an enormous number of interacting and diverse parts that are organized in a hierarchical fashion. This makes their full tractability through mathematical models quite difficult, if not impossible, in practice. Explanatory practice adjusted and expanded in response to the continued investigation of the types of complex systems found in biology and the rest of the special sciences. An increased focus on isolating and describing particular causal relations, as opposed to mathematical description and subsuming phenomena under laws, was a large part of this change.

Abstraction and idealization are still very important in causal explanation, but they don't necessarily lead to mathematical models of the system. Interventionist accounts of causation, focused on prediction and an increased control of the world, are especially suited to the sciences and will serve as my example. On an interventionist account of causation, such as the one presented in Woodward (2003), causal relations are defined in terms of relations between variables, such that intervening and changing the value of one variable changes the value of the second variable. According to the interventionist conception of causation, causal information

---

[20] I will not take a view here about whether or not these mathematical models identify, or stand in as descriptions of, the causal structure of some part of the world. The relation between causal and mathematical explanations is fraught with controversy.

[21] See Moreno et al. (2010) for a more detailed history of the impact of complexity analysis on the sciences.

provides answers to *what-if-things-had-been-different* questions. Causal explanations identify conditions under which an outcome would have been different, that is, they identify changes that can be made to manipulate the outcome of interest.

This approach proceeds by picking out and isolating a particular system or phenomenon. Assumptions are made about what the components of the system are and how they interact. This aids in the identification of relevant variables and abstraction away from irrelevant details. Then relations between variables are determined through observation or experimentation. Causal models can then be constructed, with their accuracy (and the justification of initial variable choices) usually evaluated in terms of predictive power and generality. Woodward (2010) describes three concepts that are especially important for picking out the appropriate causal variables in biological systems: stability, proportionality, and specificity. Stability refers to the degree to which a causal relationship continues to hold under a wide range of background conditions. Proportionality refers to picking variables with the right amount of detail along the particular-general continuum. Specificity is the extent to which a causal relation approximates to the ideal of one cause-one effect. All other things being equal, causal explanations are better to the extent that the cited models are detailed, complete, and accurate in the sense of identifying the full range of changes in all those factors (and only those factors) that are associated with changes in the target phenomenon (Woodward 2003, 2010).

Decomposition also plays a crucial role in causal reasoning in biology because the causal interactions making up most biological systems are too numerous or complex to investigate as a whole. Decomposition is the strategy of breaking systems into their component modular parts, investigating those parts separately, and then trying to figure out how the parts interact with each other and within the system as a whole (Simon 1962; Bechtel and Richardson 1993/2010; Levy

2014). A component part is modular if the difference it makes is independent of the difference made by other components. That is, component X's contribution is modular if it is possible to disrupt the activity of other components without affecting the contribution of X (Woodward 2003, 2013; Levy 2014).

Decomposition may proceed along structural, functional, or temporal lines and is often an iterative process because systems are almost always themselves part of larger-scale systems and made up of smaller-scale ones. Decomposition is especially important in biology because the parts of a biological system are often heterogeneous and highly organized. This is unlike some of the traditional targets of the physical sciences, such as gases and medium-size solids and fluids, where the global behavior of the parts can be averaged out and (relatively) straight-forward connections between hierarchical levels can be made.

Decompositional strategies lay at the heart of mechanistic approaches to explanation (Bechtel and Richardson 1993/2010; Machamer et al. 2000; Levy 2014; Skillings 2015). Mechanistic explanations explain by decomposing a phenomenon or process of interest and mapping out the relations of the entities and activities that produce the phenomenon or process. Mechanistic explanations typically answer *how* questions: "how does a particular phenomenon come about?", or "how is a particular function realized?".

Mechanist and interventionist approaches are seen by some to go hand-in-hand, with the details of mechanisms and complex systems gathered via an interventionist approach (Craver 2007: 160). On this view complex systems can be decomposed into mechanisms, and parts of mechanisms and their interactions can be described via experimental interventions. Moving in the other direction, the details of systems and mechanisms (in at least some cases) can be

abstracted away from and instantiated as variables in a more general interventionist causal description of a larger scale system.

I take the story as given so far to be a basic sketch of the explanatory strategies used to investigate the causal structure of the world. It, of course, leaves out a majority of the details, such as the underlying metaphysics of causality, how causal relations are experimentally verified, and how causal knowledge is used for prediction. According to this minimal picture, the strategies for discovering causal relations and generating tractable causal explanations are captured by the interaction of two approaches to simplification: the omission of irrelevant details through abstraction and idealization, and the decomposition of a system in order to highlight its underlying organization and break it up into more independently manageable subsystems. Ideally, causal explanations could be generated in-principle for all phenomena via iteration of these two strategies. When the behavior of a system is straightforward, abstract away to the salient causal factors. If the behavior is too complex, keep on decomposing it into more manageable parts until you get to a system tractable enough to where you can abstract away to the salient causal factors. Then recompose as needed. Unfortunately, this traditional line of causal reasoning runs into problems when applied to the behavior of many biological systems. The rest of the paper is concerned with showing why this is the case.

3.3 Translating the Causal Storms of the Nanoworld

The explosion of progress in biophysics, molecular biology and genetics in the last few decades has painted a new picture of the inner workings of the living world. One striking feature of the biomolecular processes that underlie all biological phenomena is that they behave in a way that is quite different from the behavior seen in the macro level processes experienced in our everyday lives. I maintain that because of this our causal explanatory habits have the potential to

seriously mislead us when thinking about the small-scale features of the biological world, which in turn might cascade into other problems elsewhere.

The difficulty is an instance of what I call the *causal translation problem*. It refers to situations where strategies of causal reasoning useful in one domain or scale are transferred to another domain or scale where they are not appropriate and/or applicable. This may be because of the simplifying assumptions used to make each domain tractable are not easy to translate between. Or, it may be that the systems behave too differently, making it difficult to integrate descriptions or explanations across scales.[22] I will further illustrate the *causal translation problem* by means of a contrastive example.

Consider two processes: the time-keeping of a mechanical clock and the synthesis of a protein by a ribosomal complex. One way to explain how they work is look at how their parts interact. What are the causal interactions of the parts that make up each process like? How should they be described? The workings of a mechanical clock are likely more familiar than those of a ribosomal complex. The mainspring stores mechanical energy through torsion. The interaction between the mainspring and the balance wheel via the escapement mechanism turn the gears at a fixed rate. These gears are attached to the rotating hands of the clock face, which indicate the time.

We can say a few more things about the mechanism of a watch. It is robust process relative to a wide variety of background conditions, including many changes in size, temperature, gravity, light etc. The parts can be made out of a wide variety of materials. Furthermore, changes

[22] This is similar to the problem Batterman calls "the tyranny of scales" (Batterman 2013).

in these variables will change very little about the way in which the process unfolds. There will likely be miniscule changes in the friction and elastic stresses where the parts interact—as a result of different metal hardness or strength, for example--but to a great extent all of the parts will interact in approximately the same way across a wide spectrum of conditions.

Protein synthesis is the process by which a ribosome constructs a new protein from amino acids in an order directed by a mRNA. A ribosomal complex has parts, but the parts don't stand in the same relations as the parts of a mechanical clock. Imagine we are studying a population of 100 identical ribosomes that are all synthesizing the same type of protein. When we compare what each ribosome did in the synthesis process we will almost certainly see a large amount of variation across runs. To understand why this is, we have to take a detour through molecular and cellular biology.

3.3.1 The World Inside the Cell

Protein synthesis is a well understood cellular mechanism. It is an incredibly reliable process. The error rate is on the order of 1 in 10,000. This regularity in production is one reason why ribosomes are thought of as tiny molecular machines. Zooming in to the level of individual ribosomes and individual cycles of protein synthesis reveals a much different world compared to what we are used to seeing in macroscopic phenomena. Let's look at what the world inside the cell is like.[23]

---

[23] In this section I draw on recent work in cellular biology, molecular biology, biochemistry and biophysics, especially Peter Moore's paper "How should we think about the ribosome?" (2012), Dean Astumian and Peter Hänggi's paper "Brownian Motors" (2002), and Peter Hoffman's book *Life's Ratchet* (2012). Additional issues are explored in Skillings (2015) "Mechanistic explanation of biological processes".

The interior of a cell is a crowded place, packed full of proteins, DNA, RNA, lipids and various other organic molecules. The scale at which these molecules exist is measured in nanometers—billionths of a meter. Physical interactions inside the cell are remarkably different from physical interactions at our scale because the relevant physical forces are quite different at each level. As volumes become small, forces associated with mass become less important and surface interactions start to dominate. Forces that are important at the macroscale, such as gravity and inertia, become irrelevant at the nanoscale. Likewise, forces that dominate at the nanoscale, such as friction, electromagnetic attraction and thermal forces, are much less important at the macroscale. For example: in a baseball game, inertia (when the bat hits the ball) and gravity (when the ball comes back down) dominate. Typically, the ball does not stick to the bat. In a game of nanobaseball, however, inertia and gravity would be unimportant, as the ball would have a mass of next to nothing. But the relatively large surface area compared with the tiny mass of the nanobaseball would make it difficult to get the nanobaseball off the nanobat[24].

Another important feature of the cell at the nanoscale is the fact that it is a liquid water environment. Every molecule in a cell is constantly barraged by fast-moving water molecules. Larger molecules, like proteins, are randomly struck by a water molecule trillions of times per second. These constant collisions add up to a lot of power--relatively speaking. At body temperature a thermal noise power of about 10^{-8} W continually swirls around a protein. That force is approximately a hundred million times greater than the force that a protein like a ribosome can generate. These thermal forces, coupled with the electrical charge, cause larger molecules to vibrate, move about, and constantly and spontaneously rearrange themselves.

[24] This example is based on one used in Hoffman (2012)

Hoffman calls this chaotic environment the *molecular storm*. For molecules, moving in a directed way under their own power would be more difficult than walking through a tornado is for us. Unlike a macrolevel storm though, the molecular storm has no directionality. Every water molecule collision comes from a random direction.

The nanoscale is special in one final way. Only at the nanoscale are many types of energy--such as elastic, mechanical, electrostatic, chemical, and thermal—at the same magnitude. This means that an appropriately structured molecule can spontaneously convert different types of energy into one another. A molecule like a ribosome can go through substantial fluctuations in energy as thermal energy randomly waxes and wanes. When the thermal energy randomly crosses a threshold, the ribosome can use it to convert, for example, chemical energy into electrical or directed kinetic energy. At smaller scales, such as at the scale of atoms or subatomic particles, binding energies are too large to be subject to thermal energy fluctuations. By contrast, at scales much larger than a nanometer, it is the mechanical and electrical energies that are too high to be subject to thermal fluctuations.

Let's now return to the process of protein synthesis. Because of the nature of events at the nanoscale, the parts of a biochemical mechanism don't interact in the same way as the parts of a machine mechanism at the macro-level. Ribosomes are the prototypical example of a group known as Brownian machines, which are now thought to be ubiquitous at the cellular level (Dashti et al. 2014). Brownian machines are macromolecular complexes buffeted by the random motions of molecules in the environment, and they are capable of exploiting these thermal motions to do work. 99.99% of the activity in protein elongation is random fluctuation that has nothing to do with function, and most of the progress through the protein elongation cycle is driven by changes in relative free energies coming from conformational state changes and the

presence of protein factors that promote some states over others (Moore 2012). This means that the ribosome is randomly pushed into different conformational states by external forces, rather than moving in sequence according to the activity and interactions of its parts.[25] A protein gets synthesized as the ribosome is pushed around and randomly visits functional conformations. Given enough time, the ribosome will eventually cycle through all of the functional states that are required to synthesize a protein. Moore (2012) states that: "[A]ll the functionally significant movements of the ribosome, both internal and external, are biased random walks, and it is most unlikely that any given ribosome will ever do exactly the same thing twice as it elongates some polypeptide".

If we were to trace all of the movements of a ribosome during protein synthesis, it is unlikely that we would trace the same path twice when comparing across events. We won't observe a regular succession of events across instances like we would if we were observing the inner workings of a mechanical watch. This is an important difference in the causal behaviors between different scales in that it conflicts with a central intuition about the behavior of causal chains. "Obviously, as a general rule, more distal causal relationships with many intermediate links will be less stable than the individual links themselves" (Woodward 2010: 294). The search for the stable relationships that are in turn held up as the important causal factors will tend to drive in a "reductive" direction for more fine-grained details, "a search for the more proximate

[25] Biomolecules at the nanoscale are a perfect example appear to be a perfect example of what Wimsatt calls "causal thickets", which are systems characterized by large degree of cross-level interaction (Wimsatt 1994). According to Wimsatt, causal thickets pop up when the system of interest is at the interface of two scales. This makes it difficult to give a single causal description of the system's behavior. The two scales here being that of the protein complex and that of the water molecules whose aggregate behavior is abstracted as thermal energy at the level of the macromolecule.

relationships that mediate distal relationships". Woodward admits that this isn't always the case, but the example above suggests that there are large groups of important processes where this doesn't hold. In the case of protein synthesis the entire process is very stable, and the visitation of intermediate functional conformations perhaps even more so, but the even more proximate movements between functional conformations are stochastic and not stably ordered at all.

There is no regular ribosome movement, and this isn't because the world is messy. It is not the case that under ideal conditions we would get identical protein synthesis events. That is how the watch works. Under ideal conditions and an identical starting position the watch behaves the same way every time. Protein synthesis is itself stochastic, there are no ideal conditions under which we could remove variation. A ribosome that didn't have its movement dominated by external stochastic forces would not do anything at all, because that environment would only exist at 0 Kelvin, an environment where there is no biological activity. If you remove the source of the variation, the process itself fails. We could never predict the precise movements of a particular ribosome given past observations of other ribosomes, or even past observations of that exact ribosome.

An abstract functional explanation tells us a lot about protein synthesis, but it leaves out the details of how those functions are connected[26]. Moore makes a similar claim. The central thrust of his paper is the claim that if an explanation "of elongation is not to be totally misleading, it must depict the endless series of meaningless, thermally driven, conformational fluctuations that separate one functionally significant event from the next" (Moore 2012). He recommends that we "come to think of the ribosome as the dynamic, constantly varying structure

[26] Refer back to chapter 1 and (Skillings 2015) for further discussion of this point.

it is and always was" and move away from quasi-mechanical explanations for the properties of ribosomes and "instead seek to understand them in terms of the particle's conformational energy landscape" (Moore 2012).

A functional explanation of protein synthesis involves the determination of a complex mechanism, much like the explanation of a mechanical clock. Each mechanism is linear, and the parts interact in a direct and fixed way. But if we look at the process by which an actual protein is produced, the similarities between a watch and a ribosome begin to evaporate. The functional explanation covers up the important ways in which a ribosome is different from a watch[27]. There is no productive continuity between stages, where earlier functional events directly produce later functional events. Rather, protein synthesis proceeds stochastically. Watches reliably tell time, and ribosomes reliably produce proteins, but the relations between the temporal stages of the mechanisms are much different. A watch has a fixed structure and operates in a fixed sequence. The behavior of molecular machines is explained in terms of their free-energy landscape. The parts of the ribosome are constantly in motion and randomly changing their relations to each other due to forces that dominate at the molecular level. These ongoing interactions are what dynamically stabilize the ribosome even as it flows constantly from conformation to conformation in a biased random walk. Furthermore, it is this constant change that leads to the functionally important states needed for protein synthesis. Knowing this is critical for understanding how the ribosome synthesizes proteins.

[27] This may be a general downside of cashing out explanations solely in terms of functions. Not knowing the details of how the function is realized limits our ability to carry out interventions, predict the effects of perturbations, and generally answer *what-if-things-had-been-different* questions.

Biological systems at the macroscopic scale, such as organs and tissues within an organism, tend to behave more like machines. Parts interact by pushing and pulling on each, because at that level the forces that generate Brownian motion are ineffective. The fact that many biological systems can be approximated in such a way made the organism-as-a-machine assumption a very powerful conceptual approach. Upon closer investigation, and especially at smaller scales, the machine metaphor rapidly shakes apart. Ribosomes and other molecular machines don't behave like macroscopic mechanical devices. Using them as analogies for one another is misleading. Any conclusions about one system will probably be inaccurate if they are based on an analogy of how the system at the other scale acts[28]. The causal behaviors of these systems do not directly translate back and forth.

3.3.2 Re-imagining the physical basis of the mind

Increased awareness of differences in the behavior of biological systems could also change the shape of other persistent problems outside of biology. For example, Peter Godfrey-Smith argues that the behavior of metabolic and cellular processes forces us to rethink the so-called "explanatory gap"[29] problem, which hinges on intuitions about how subjective experience might arise from physical systems (Godfrey-Smith 2016). Conceivability arguments against materialism depend on intuitions about the separability of the mental and the physical.[30] This argument states that it seems like we can conceive of an exact physical duplicate that is the same

---

[28] This point is echoed in (Astumian et al. 2016): "A complete understanding of the mechanism of a molecular motor can be gained only by viewing the chemo-mechanical cycles holistically. Experiments on parts of the cycle can fill in the details, but over-interpretation often leads to erroneous conclusions, especially when analogies are drawn with macroscopic elements".
[29] See Levine J. 1983. Materialsim and qualia: the explanatory gap. Pacific Philosophical Quarterly LXIV. October:354-361.
[30] For example: Nagel, Kripke, and Chalmers

in every way to that of an ordinary human excepting for it not having any subjective experience. This is supposed to show that the mental and physical are separable, and hence the failure of materialism. This argument depends on some crucial intuitions about the physical behavior of biological systems. Intuitions that probably don't match the actual behavior of those systems. Godfrey-Smith writes:

> In us, the material basis for mental activity is tied to cells and metabolism. When we look at what's actually going on in our bodies and brains, we find that many of the imaginatively familiar features of the physical are not present. Many of the features of the physical that strike our imagination in a way that seems un-mental are not present. And it is difficult to imagine the crucial processes at all, hard to get any sort of intuitive handle on what they are capable of. Arguments against materialism based on conceivability rely on the trustworthiness of intuitions about what the particular physical processes inside us can produce. Once we see what those physical processes are actually like, the trustworthiness of the crucial intuitions is much reduced. (Godfrey-Smith 2015)

Redescribing this in my terminology: conceivability arguments against materialism require dependable intuitions about the behavior of small-scale biological processes and how they are related to higher-level processes such as perception, consciousness and subjective experience. But, our causal reasoning habits have been formed through experience with the macro-world and likely don't translate very well to thinking through complex systems and causal interactions at the scale of cells, where things behave much differently. The onus is on those arguing for the trustworthiness of our intuitions to show that the *causal translation problem* has been overcome.

3.4 Interpreting Complexity in Living Systems

I now turn to what I call the *causal interpretation problem.* Many biological systems are complicated in a way that goes beyond the complexity seen in the systems studied by the physical sciences. Biological systems have many features absent in non-living complex systems: hierarchical organization, long-term sustainability, historicity, functional diversity, adaptivity and agency. Living systems exist in nested and interconnected levels which, being somewhat self-organized in their local dynamics, depend globally upon each other (Moreno et al. 2009, Moreno and Mossio 2015). Living systems are not just more of the same, they are qualitatively different. My claim (echoing others[31]) is that living systems present a new type of causal regime. The *causal interpretation problem* stems from the difficulty in producing causal descriptions of these systems that are both adequate and tractable at the same time.[32]

Here are three sources of causal complexity:

(1) a large number of interactions in a system,

(2) differences in the kinds of interactions between parts of the system,

(3) constraints placed by the system as a whole on the interactions of the parts of the system.

Interpreting the causal structure of a system becomes increasingly difficult as these three sources of complexity interact. Many biological systems exemplify this through their high number of components, interacting in non-linear ways while constrained or channeled by the global properties of the system. This introduces a holism in the system (Mitchell 2009; Moreno

---

[31] For example: Bechtel 2007; Mossio et al. 2013; Mossio and Bich 2014.

[32] The *causal interpretation problem* shares similarities with Sara Green's diagnosis of the difficulties in explicating the structures and capacities of complex system. Where I collapse all these difficulties under one label, she divides them into what she calls the *synchronic* and *diachronic underdetermination problems* (Green 2015).

et al. 2009), in the sense that decomposition of the system into its parts is either not possible, or comes with a loss in regards to understanding the behavior of the system as a whole because the behavior of the system is not an aggregation of the behavior of its parts.

If decomposability and modularity are required for investigating the causal contribution of individual parts to a system's behavior (such as Woodward 2003), then an interpretation of the causal behavior of a complex system becomes problematic. A component may have a causal role when the system is in an undisrupted state, but have a different, or no, causal role when the system is in a perturbed state, such as when the system is manipulated and decomposed for experimental purposes (Mitchell 2009).

Self-maintaining and self-producing systems, such as cells and some regulatory networks, are particularly complex and often the most resistant to decompositional strategies. Many feature dynamical causal loops, such that, at least one system-level property (itself the product of the interactions of the parts) is causally necessary for the maintenance of the loop (Bechtel and Richardson (1993/2010; Mitchell 2009). When the interactions between component parts change their structure and dynamics when the system is decomposed into sub-systems for isolated study (a practical necessity), then this can affect inferences about their stability and characteristic function when put back together (Simon 1962; Moreno et al. 2009). This makes it difficult to understand the causal interactions of the system because you can't study the parts in isolation without disrupting the workings of the entire system.

Braun and Marom (2015) highlight additional features that add to the difficulty of decomposing biological systems. First, nearly identical microscopic configurations can give rise to different macroscale dynamics and functions, and different microscopic configurations can

give rise to the same macroscale dynamics[33]. There is a many-to-many degeneracy in system function. This makes an interventionist approach to causal explanation a lot less powerful. In systems with many-to-many degeneracy, seeing no effect when a variable is manipulated doesn't mean that the variable isn't causally contributing to the behavior of the system. Other parts of the system could be compensating[34]. Second, there is no time scale separation between the levels of organization and so systems can't be decomposed according to the speed of their interactions. Smaller processes almost always occur at faster rates than larger-scale processes. The difference in interaction speeds between scales is one of the most fruitful decompositional approaches for organizing the behavior of physical sciences. It is less helpful in biological systems.

The complexity and robustness of holistic biological systems frustrates attempts to simplify them and make them tractable while maintaining a high degree of accuracy or usefulness. The *causal interpretation problem* is meant to capture this fact. A complete description of all of the salient causal factors and their interactions inside a complex holistic system would almost certainly be too unwieldy to be useful. Abstracting away from too many of the causal factors in these systems can make the resulting description useless for predicting the behavior of the system or intervening on it. Possibly the most pernicious aspect of this problem comes from the difficulty of deciphering whether or not one has accurately captured the causal relations between the parts of such complex holistic systems at all.

Systems biology is one new field focused on understanding the complexity of holistic

---

[33] Lewontin (1970, 1974) makes this same point when talking about many-to-many genotype-phenotype mapping.

[34] See reviews by Nijhout (2002), Kitano (2004), Wagner (2005), Bateson and Gluckman (2012), Whitacre (2012), and chapter 2 for discussion of robustness in biological systems.

systems (Kitano 2007, Green 2015). This approach attempts to determine how systems-level properties, such as functional properties, arise by non-linear interactions between the component parts.[35] One way to do this is by finding common network architectures in the organization of complex systems that help explain why the systems behave the way that they do. There has been some success in finding some network motifs, such as the small-world networks and bow-tie networks common to robust systems (Kitano 2004; Whitacre 2012; Green 2015).

3.5 Causal Reasoning in Biology Redux

Abstraction and decomposition are powerful and probably indispensable conceptual tools for reducing the complexity of natural systems to a degree where tractable descriptions and explanations can be given. But, in the end, I suggest they are best viewed as heuristics (Wimsatt 2007). The heuristic viewpoint shows why it is important to pay attention to both how our conceptual tools resolve problems and when and where they can break down. The *causal translation problem* and the *causal interpretation problem* capture two types of scenarios where abstraction and decomposition have a greater tendency to break down. Causal explanations constructed in domain-specific ways may be adequate for meeting prediction or manipulation goals in that domain (or at least adequate stand-ins for explanations in-progress), but they can be very misleading in other contexts if not applied with care. Decomposing phenomena as if they are machines can be an efficient method for understanding macro-scale phenomena, but is usually counterproductive for understanding phenomena at the nano-scale. Paying close attention to biological practice and theorizing is an important preventative measure for avoiding the

[35] Refer back to Chapter 2 for further discussion of systems biology approaches, especially in the context of the importance of robustness for the persistence of complex systems.

problems with adequately translating and interpreting causal relations in biology.

INTERMEZZO

The dissertation so far has focused on some of the strategies that we use to try and understand the biological world. It has argued that the complexity of living systems challenges the adequacy of traditional approaches to scientific explanation because currently dominant theories of scientific explanation don't adequately track the explanations given in the biological sciences. My project has been then to show how the details of the biological world motivates an expansion of our considered accounts of scientific explanation. I have also shown how the empirical investigation of living things give us reason to be doubtful of our casual intuitions and our causal reasoning habits.

When considering complex phenomena, or phenomena at scales vastly different than our own, it is important to keep in mind how our intuitions might lead us astray. Our causal reasoning habits have been formed in response to a world experienced at a medium scale. Things behave in an unusual way at the microscopic and nano scales. Getting access to those worlds has challenged our understanding of how living systems operate.

The next chapter moves away from broad questions about explanation in biology to zoom in on an active debate within biology. Which entities are biological individuals, and how should they be characterized? A fundamental question within biology, if there are any. Though the material found in this chapter may appear to be disparate at first glance, two background themes that have emerged as running through the project so far are also to be found here. The first is a question: what are the appropriate conceptual and theoretical tools for understanding the living world? The first three chapters look at broad frameworks (mechanistic explanation, robustness and system-level explanation, causal reasoning), the next chapter focuses more narrowly on

theories of biological individuality. The second theme is a focus on how the empirical investigation of the microscopic and nano worlds has put into question the assumptions of dominant theories used to understand the biological world.

In the first three chapters I used details about the inner workings of cells to argue for a change in how we think about living systems more generally. In the next chapter I investigate whether the details of microbial interactions should change how we think about biological individuals, including how we think about ourselves. In these ways chapter four is a continuation of the first three chapters. It is a case study where the broader themes are demonstrated through wading in on an ongoing debate about how to understand the evolution and interactions of living things.

CHAPTER 4: Holobionts and the Ecology of Organisms – Multi-Species Communities or Integrated Individuals?[36]

4.1 Introduction

Multicellular organisms have been engaged in symbiotic relationships with microorganisms throughout their evolutionary history (Moran 2006). It was long thought that all macroorganisms are routinely colonized by a large number of microorganisms, but the details and extent of macrobe-microbe interactions remained difficult to uncover. As molecular sequencing technologies have advanced, the microbial world has been increasingly opened up to investigation by biologists. It is now widely accepted that microorganisms have always played– and still continue to play–many important roles in the lives of plants and animals (McFall-Ngai et al. 2013, Douglas and Werren 2016). Symbiotic interactions between microbes and macrobes have been documented among diverse organisms, and many researchers maintain that all macrobes engage in symbiotic interactions with microbes in natural settings (Zilber-Rosenberg and Rosenberg 2008; Bosch and McFall-Ngai 2011; Dupré and O'Malley 2009; Gordon et al. 2013; Singh et al. 2013; Booth 2014).

The term "holobiont" was coined by Lynn Margulis and used to refer to symbiotic associations that last throughout a significant portion of an organism's lifetime (Margulis 1991). The term first found wide usage in coral biology where it was defined as a coral colony and its associated photosynthetic algal symbionts and bacterial communities (Rohwer et al. 2002; Knowlton and Rohwer 2003; Stat et al. 2012). The recent influx of interest in macrobe-microbe

[36] This chapter was published as: Skillings D. 2016. Holobionts and the ecology of organisms: multi-species communities or integrated individuals? Biology & Philosophy. doi:10.1007/s10539-016-9544-0.

relations has led to a proliferation of the term "holobiont", now most often understood as a host macroorganism and *all* of its associated microbiota, including bacteria, archaea, viruses, protists, fungi, and microscopic multicellular animals such as nematodes (Zilber-Rosenberg and Rosenberg 2008; Booth 2014; Bordenstein and Theis 2015; Moran and Sloan 2015; Douglas and Werren 2016; Theis et al. 2016). Because the holobiont includes *all* associated microbiota, the interactions between holobiont partners may be harmful, beneficial or of no consequence. Relationships between partners may be coevolved or opportunistic, competitive or cooperative.

Recognition that holobionts are common in nature has led many researchers to reassess their views about various processes and concepts that are foundational in biological thinking. Dupré and O'Malley (2009) was an important early step in this direction, suggesting that microbial biology has radical implications for the future development of many areas of philosophy of biology. Philosophers are not alone here. They are joined by biologists in calling for a transformation in our thinking. "Right now, for those of us who are not evolutionary biologists, it is enough to recognize that the very foundations of biology are being shaken by both the integration of microbiology into concepts of macroevolution and the recognition that host-microbe symbioses are a major theme in biological systems" (McFall-Ngai 2016). One suggestion is that there is now a need to "upgrade fundamental theories" because holobiont systems "raise the discussion of individuality and organismality beyond its historical perspective to a level that challenges and extends current thinking" (Bordenstein and Theis 2015).

Criteria for individuating entities are of central importance in biology (Hull 1978, 1980, 1992; Buss 1987; Godfrey-Smith 2009, 2013; Clark 2011; Pradeu 2012). For example, population biologists and ecologists must be able to distinguish individuals in a population. Evolutionary biologists must be able to distinguish parents from their offspring, and one lineage

from another. Immunologists and physiologists must be able to distinguish between an individual and its environment. Organisms have long been the paradigm of individuality—a horse, a tree, a human--both within and outside of biology (Aristotle 1984). Extensive experimentation and theoretical advances in biology, especially within the last century, have changed our understanding of how individuals can and did evolve (Buss 1987; Maynard Smith and Szathmáry 1995; Michod 1999). New conceptions of individuality have helped us to understand individuality across the biological hierarchy: genes, cells, multicellular organisms, superorganismal colonies, and multi-species symbiotic communities (Dawkins 1976; Sober and Wilson 1989; Queller and Strassmann 2009).

The question motivating this chapter is: are holobionts biological individuals or communities? I start by introducing the biology of the coral holobiont as an example of the complexity and diversity of the interactions within a holobiont. I take the coral holobiont to be a good example of typical holobiont dynamics, and as such an appropriate test case for working through whether or not holobionts are biological individuals. In the following sections I expand on relevant accounts of biological individuality and claims made in favor of holobionts being biological individuals. I then consider whether holobionts meet some plausible and common criteria for either evolutionary individuality (reproduction and heritability) or organismality (functionally integrated interactors, metabolic collaboration, or cooperative low-conflict consortiums). I conclude that most holobionts share more affinities with communities than they do with individual wholes, and that, except for in rare and possibly unrealized cases, holobionts *do not* meet the criteria for being evolutionary individuals, units of selection, or organisms.

4.2 The Coral Holobiont

The evolutionary and ecological success of corals in the characteristically nutrient-poor

environments of tropical and subtropical oceans is thought to be a direct consequence of their ability to form mutually beneficial symbioses with unicellular photoautotrophic dinoflagellates in the genus *Symbiodinium*, commonly referred to as zooxanthellae (Stat et al. 2012; Lesser et al. 2013). The appearance of coral reefs in the Triassic has been attributed to the evolution of the symbiotic association between the coral host and *Symbiodinium*.

Corals are mainly colonized by free-living *Symbiodinium* from the environment, but a direct transfer of *Symbiodinium* from parent to offspring via inclusion in the egg also occurs in many brooding corals, and is occasionally seen in some spawning corals (Thompson et al. 2015). The *Symbiodinium* are endosymbionts that reside within the cells of the coral host. They fix carbon through photosynthesis, which they provide, along with other nutrients, to their host in return for host waste metabolites and protection from grazing (Stat et al. 2012; Roth 2014). This exchange of nutrients is mutually beneficial and helps the coral secrete the calcium carbonate skeletal structure that is shared by the coral colony and contributes to the formation of coral reefs. This endosymbiotic association is especially interesting because it involves two eukaryotic organisms and the genome of the symbiont is three times larger than the genome of its host (Roth 2014).

The total number of bacterial species associated with corals remain largely unknown, but recent estimates put the number between 3000-6000 species (Stat et al. 2012). At least some of the coral-associated prokaryotes are beneficial to the coral host. Cyanobacteria provide nutrition through nitrogen fixation (Lesser et al., 2004), and bacteria residing in the coral's exterior mucous layer act as a first line of defense against pathogens by producing antimicrobial compounds and occupying space (Stat et al. 2012). Most coral-associated fungi and viruses are thought to be parasitic (Golubic et al. 2005), but there may be exceptions. There is evidence that

some endolithic fungi residing in the coral skeleton could be converting nitrate and nitrite to ammonia which could enable fixed nitrogen to cycle within the coral holobiont, and some of the bacteriophages might be helping to beneficially regulate associated bacterial communities (Wegley et. al. 2007). With rare exceptions, the microorganismal component of the coral holobiont is obtained from the environment (Apprill et al. 2009).

There is a substantial amount of genetic diversity within *Symbiodinium* and evidence that some of it is reflected as functional diversity. This is important because corals can harbor more than one type of *Symbiodinium* at a time, and are known to shuffle and switch out their *Symbiodinium* types. The coral holobiont is very sensitive to changes in ocean temperature and lives close to their upper thermal tolerance limit. A prolonged temperature increase of as little as 1 ˚C causes stress and can lead to coral bleaching, the forceful expulsion of the *Symbiodinium* by the coral host. The coral holobiont is a dynamic system with members fluctuating on a daily basis depending on the environmental conditions and life-cycle requirements (Thompson et al. 2015).

For example, corals associated with clade C *Symbiodinium* usually perform better than corals with clade D *Symbiodinium* in normal conditions, with clade C-infected juveniles growing two to three times as fast as those infected with clade D (Lesser et al. 2013; Hume et al. 2016). But in hot environments where coral bleaching is common, corals with Clade D *Symbiodinium* do better. One reason is that clade D appears to be more heat-tolerant than Clade C, and may rapidly increase the heat stress tolerance of corals (Lesser et al. 2013; Thompson et al. 2015; Hume et al. 2016). This can lead to short-term benefits during periods of thermal stress, but corals dominated by clade D *Symbiodinium* show significantly decreased growth and reproduction in the long term. This suggests that the *Symbiodinium* switch between being

mutualists and parasites of the host depending on current environmental conditions.

It has also been suggested that things are not what they seem for the *Symbiodinium* either. The coral host might be benefiting at the expense of their microbes by capturing and controlling its algal symbionts (Garcia and Gerardo 2014). Wooldridge (2010) suggests that "the coral host exerts a 'controlled parasitism' over its algal symbionts that is akin to an enforced domestication arrangement" with "…the coral host as an active 'farmer' of the energy-rich photoassimilates from its captive symbionts."

Which lineages come together to make up a coral holobiont is strongly contingent on environmental conditions and cooperation and conflict between different possible partners. This flexibility in lineage composition allows the coral to respond to the abiotic environment in ways that it wouldn't be able to on its own, and is likely important for the success of corals. Understanding the mechanisms and relations that have evolved between the coral and its symbiotic partners is crucial for understanding coral biology.

4.3 Controversy about the Status of Holobionts

Several biologists and philosophers have claimed that holobionts, or similar multi-lineage assemblages of macrobes and microbes, constitute at least one level of organization at which natural selection acts. Biologists Zilber-Rosenberg and Rosenberg were the first to articulate what they call the ''hologenome theory of evolution,'' which they see as an alternative to ''currently accepted dogma,'' according to which the units of selection are individual organisms as traditionally conceived (2008: 731). The hologenome is a collective unit made up of all of the host and microbial genomes of the holobiont. They write, "In the hologenome theory of evolution, we suggest that the holobiont… with its hologenome, acting in consortium, should be

considered a unit of selection in evolution…'' (2008: 723). They are not the only theorists to have made such claims:

- "Therefore the holobiont, i.e. the host including all symbionts, should be regarded as the unit of selection as the association between host and symbionts may affect the fitness of the holobiont depending on the environment" (Feldhaar 2011).
- "The hologenome theory of evolution considers the dynamic holobiont as a single unit for natural selection and provides a more accommodating view of evolution blending Darwinism and Lamarkism" (Singh et al. 2013)
- "[A]n organism's genetics and fitness are inclusive of its microbiome" (Brucker and Bordenstein 2014).
- "Thus, the holobiont, with its integrated community of species, becomes a unit of natural selection whose evolutionary mechanisms are largely unexplored" (Gilbert 2014).
- "The hologenome concept is a holistic view of genetics in which animals and plants are polygenomic entities. Thus, variation in the hologenome can lead to variation in phenotypes upon which natural selection or genetic drift can operate" (Bordenstein and Theis 2015).

Without referring to holobionts explicitly, philosophers Dupré and O'Malley[37] endorse a similar view about the fundamental entities that are operated on by natural selection: "…complex systems involving the collaboration of many highly diverse lineage forming entities. This sort of

[37] O'Malley has since moved away from this view, recently stating that natural selection probably does not act at the collective level in multilineal systems, of which holobionts are one kind (O'Malley 2016).

interactor, we also suggest, is the most fundamental unit of selection." (Dupré and O'Malley 2009). Ereshefsky and Pedroso (2013, 2015) use biofilms as a case study to defend the position that multi-species consortiums can be units of selection.

Other authors have been critical of the claim that holobionts are important units of selection. Moran and Sloan (2015) state that "While biologists would agree that microorganisms have important roles in host evolution, this statement is a far cry from the claim that they are fused with hosts to form the primary units of selection, or that hosts and microorganisms provide different portions of a unified genome" and that "…some observations that superficially appear to support the concept of the hologenome have spawned confusion about real biological issues." Douglas and Werren (2016) state that "…it is highly unlikely that the entire microbiome will evolve as a "holobiont" with its host" and that "…the hologenome concept is unhelpful to the study of host interactions with resident microorganisms…"

Pushing back against recent criticism, Theis et al. (2016) have backed off a bit from claims that holobionts are always units of selection, but still consider the holobiont to be a level at which selection acts. They also reiterate that shifts in the microbial community are akin to changes in allele frequency in the host genome, suggesting that the hologenome is a single unit upon which selection acts.

4.4 Biological Individuality

I will start with some general remarks about biological individuality, before moving on to considerations about whether or not holobionts are biological individuals. Accounts of biological individuality tend to be clustered around a few different investigatory projects. These accounts are not exclusive, a particular biological entity may be an individual of more than one type.

Genealogical individuals are lineages such as species and phylogenetic taxa (Hull 1978). They are the units that can evolve. Evolutionary individuals are the individuals of natural selection (Sober and Wilson 1998; Ereshefsky and Pedroso 2015; Clarke 2016). They are the units upon which natural selection operates and members of a population that has the capacity to evolve.

The traditional target of accounts of biological individuality is the organism, the phenomenologically discrete living entities inhabiting the world around us. I contrast organisms with evolutionary individuals in that the defining criteria of organismal individuality are not restricted to purely evolutionary considerations. As a first pass, organisms are bounded individuals that are functionally or metabolically integrated. They are systems with mutually dependent components that work together to maintain the system's structure or developmental trajectory (e.g., Pradeu 2010; Godfrey-Smith 2013). There are numerous accounts of organismality, many of which don't agree[38].

One other approach to organismality is one that focuses on physiological individuality. This is a family of views that have developed somewhat independently of evolutionary views about individuality. Of these, the immunological account advocated by Pradeu (2012) is especially promising. On this account, the boundaries of physiological individuals (organisms) are established by the immune system of the host. Other nearby physiological views rest on the fact that symbionts are either critical for host development, or make something of the host work, that is, realize or help realize an important physiological function (See, for example, Bocci 1992;

[38] See Clarke (2011) for a thorough survey of accounts of organismality and individuality.

Berg 1996; Xu and Gordon 2003)[39]. This line of thinking has been important for understanding the boundaries, health and development of macroorganisms, especially large and complex vertebrates like us.

I will not discuss the relation between holobionts and physiological individuality for two reasons. First, I am concerned with starting from a more general analysis that doesn't privilege macroorganisms or index claims of individuality to the host. I take physiological individuality to be host-centric, subordinating microbe individuality to functional, immunological, and developmental considerations regarding the host. My focus is: how might higher level individuality emerge out of the general interactions between macrobes and microbes? Second, much of the controversy and debate surrounding holobionts has focused on whether holobionts ought to be understood as units of selection or communities, and the present paper is an attempt to take side in that controversy. Because of these reasons I will limit myself to evolutionary individuality and accounts of organismality that focus on functional integration of a collaborative or codependent nature.

4.5 Are Holobionts Evolutionary Individuals?

Evolutionary individuals are entities defined in terms of natural selection: they vary among each other, their variability causes variations in fitness, and that variation and fitness effect is heritable (Lewontin 1970; Godfrey-Smith 2009; Ereshefsky and Pedroso 2015; Clarke 2016). The Darwinian population framework is one way to make evolutionary individuality more precise. This framework has its roots in the account of natural selection articulated in Lewontin

[39] I thank Thomas Pradeu for emphazing this point, as well as pointing out the gap in my treatment of physiological individuality in a previous draft.

(1970), and gets its name from Godfrey-Smith's (2009) extended update of that account. Building from Lewontin's three criteria of *variation*, *heredity* and *differences in reproductive success*, a *Darwinian population* is defined as "a collection of causally connected individual things in which there is variation in character, which leads to differences in reproductive output (differences in how much or how quickly individuals reproduce), and which is inherited to some extent" (2009: 39). A member of such a population is a *Darwinian individual* (Godfrey-Smith 2009: 40). Darwinian individuals are units of selection, and as such are "the loci of causal action for the process of selection" (Booth 2014: 664). If holobionts are to be evolutionary individuals on an account like this, then they must (a) be able to reproduce at the level of the holobiont, and (b) there must be heritable differences at the level of the holobiont. I will argue in the next section that holobionts fulfill neither of these criteria.

4.5.1 Holobiont Reproduction

Godfrey-Smith (2015) makes further distinctions that are helpful for understanding the generation of new entities. He starts with the notion of recurring structures, which may either be reproducing or reconstructed. Reproducing things form parent-offspring lineages, whereas reconstructed ones do not. Reconstructed objects include organs and enzymes. Reproduction can be collective, simple, or scaffolded (Godfrey-Smith 2009). Collective reproducers include multicellular organisms. Simple reproducers, such as bacteria, can give rise to more things like themselves. Scaffolded reproducers rely on external machinery for their reproduction. Examples include genes and viruses.

An alternative analysis of reproduction, Griesemer's (2000, 2014, 2016) "reproducer" account, shares many similarities with the Darwinian individual account. It differs in that he emphasizes the necessity of material overlap between generations. "Reproducers are entities with

the capacity to multiply because offspring bear relations of 'material overlap' with their parents" (Griesemer 2016). What makes reproduction different than recurrence or mere production is the conveyance of developmental cycles linked together in a lineage.

Griesemer and Godfrey-Smith agree that reproduction requires the formation of lineages. Their disagreement centers on the necessity of material transfer between generations and how to parse the divisions between scaffolded and collective reproducers (Griesemer 2016). For Griesemer, nearly all organisms are scaffolded in some way because they depend on some aspect of the environment for either their development or reproduction.

Questions about whether holobionts are evolutionary individuals—i.e., natural selection operates at the level of the holobiont—are intimately related to questions about holobiont recurrence. Specifically, is the generation of a new holobiont the result of reproduction or reconstruction, which is to ask: do holobionts reproduce as a whole such that that there are holobiont-level lineages? The answer depends primarily on how the microbial symbionts are transmitted.

The transmission routes by which microbial symbionts move from host to host vary considerably and are usually divided into two categories (Bright and Bulgherisi 2010). Vertical transmission is the direct transfer of symbionts from the host parent(s) to their offspring. Horizontally transmitted symbionts are acquired from other non-parental hosts or from free-living population in the environment. The majority of microbial associations of multicellular animals and plants are thought to be horizontally transmitted (Moran and Sloan 2015). Cases of vertical transmission are often not obligate, that is, even though the symbionts can be, or even often are, transmitted vertically, they can also be obtained horizontally. Obligate vertical

transmission is thought to be rare, and may reliably occur only in cases of endosymbiosis. Many cases of vertical transmission are really part of a mixed-mode of transmission. Though it appears to be the case that rates of vertical transmission are exceedingly low relative to the rates of horizontal transmission, it is perhaps the case those microbes that are transmitted vertically play a relatively larger role in the lives of their hosts[40].

Few, if any, holobionts as individual units are reproducers because strict vertical inheritance is rare. If the set of lineages that make up the holobiont varies within and between host generations, then the holobiont cannot be a coherent unit of selection. More select partnerships between hosts and individual symbionts do in rare cases meet the criteria for reproduction, such as eukaryotes and their mitochondria or corals that pass along *Symbiodinium* in their eggs. High partner fidelity is fleeting without strict vertical transmission. If the component lineages can all go their separate ways between reproductive events and reassemble at a later time in at least a semi-random fashion, then there are no higher-level lineage connecting generations of holobionts.

The only way to draw something approximating parent-offspring relations between holobionts without vertical transmission of *all* the component lineages is to privilege one of the partners. This is nearly always the host, as it is the largest and likely the longest living part of a holobiont. Privileging the host is nearly inevitable, as holobionts are defined by picking out particular hosts with all of their associated microbes.

But what if we privileged one of the other collaborating lineages? Consider the human +

---

[40] I thank Thomas Pradeu for emphasizing this point.

gut-microbiota holobiont with a different emphasis. A bacterium lives inside a doctor who hasn't been particularly careful about sanitation. The doctor goes into work and delivers an unrelated baby. The gut bacterium replicates and one of the offspring bacteria quickly colonizes the infant. A new holobiont is assembled in the collaboration of the human lineage and doctor's bacterial lineage. We can now pick out a new parent-offspring relation between the doctor holobiont and the baby holobiont. From a lineage-neutral perspective at the holobiont level, this is no stranger than saying that the parent-offspring relation is between the mother holobiont and the baby holobiont.

More sensibly, we can say that there is no fact of the matter about what constitutes the parent-offspring relation between host-microbe associations without vertical transmission. Perhaps even better: there are no parent-offspring relations between holobionts in these cases. The related concepts of parenthood and reproduction have simply been stretched too far, and most holobionts are marginal reproducers at best (Godfrey-Smith 2009, 2011, 2013; Booth 2014). In cases of horizontal transmission, the host-microbe associations recur in each generation, but they do not reproduce as a unit and do not form lineages (Godfrey-Smith 2012; Booth 2014). A particular host-microbe association might even be obligatory for the reproduction of one of the partners. But the reproductive events wouldn't be holobiont reproduction, rather they would be co-dependent scaffolded reproduction of partner lineages. That is to say, they are not evolutionary individuals on this account.

4.5.2 Holobiont Heritability and Holobiont Lineages

Holobiont reproduction with vertical transmission in itself is not sufficient for evolutionary individuality. Horizontal transmission or symbiont exchange during the host's lifetime can disrupt heritability. Even hypothetical holobionts where all of the microbial

symbionts are vertically transmitted during host reproduction may not be evolutionary individuals. The second important consideration regarding selection at the holobiont level is whether or not there is partner fidelity: a stable association of host and symbionts across multiple host generations. High partner fidelity is a prerequisite for evolutionary individuality because the holobiont can only evolve as a unit if the host and its symbionts co-occur across multiple host generations. Only holobionts with both a high degree of vertical transmission and high partner fidelity will meet the criteria for evolutionary individuality.

Partner fidelity is expected to be highest when there is obligate vertical transmission, though high partner fidelity might also be possible in holobiont systems with horizontal transmission where the hosts provide their offspring with symbionts and where specificity is high (Douglas and Werren 2016). Partner fidelity is often imposed by vertical transmission because the microbial partners will have a strong selective interest in the reproductive fitness of the host when their fitness is tied to the reproductive success of the host. But partner fidelity and vertical transmission can come apart. Partner fidelity is expected to be lower when there are no obligate dependencies, or when the obligate dependencies can be supplied by many different symbiont partners. If partner lineages can jump ship, and horizontally transfer to other hosts, then the different parts of a holobiont aren't locked into a common fate. This leads to an expectation of increased conflict between the members of the holobiont as they "pursue their own goals"; namely, selection for increased replication of one's own lineage at the expense of the success of the multi-lineage holobiont. As conflicts of interests among partners increase (e.g., due to weak partner fidelity), then the holobiont is undermined as a higher-level unit of selection.

This is especially apparent if we consider a hypothetical coral holobiont where all partners have long life-spans. For example, coral A and dinoflagellate symbiont C are distinct

lineages that interact to form coral holobiont α. Coral B and dinoflagellate symbiont D are distinct lineages that interact to form coral holobiont β. If the fates of those lineages are tied to the success of the holobiont that they help produce, then we have some notion of alignment of fitness. But if the collaborating lineages are independent of each other, it is possible that lineages C and D could switch partners. In that case A and D would interact to form a new holobiont, γ, and B and C would interact to form the new holobiont δ. Holobionts γ and δ survive, and lineages A, B, C and D eventually reproduce and make it into the next generation of holobionts.

The extinction of coral holobionts α and β did not prevent the proliferation of the lineages that produced them. The death of α and β might have fitness consequences for the individual lineages, but it needn't. In this example, the overall success of the holobiont(s) will have fitness consequences for the individual lineages that make them up, but the dissolution of any particular holobiont—because of either partner death or partner switching—needn't necessarily have fitness consequences for the individual lineages. Holobionts γ and δ could reproduce as whole units and still wouldn't be units of selection without high partner fidelity over the life of the partner lineages. A high degree of symbiont swapping will undermine selection at the level of the holobiont because horizontal swapping continually dissolves and creates individual holobionts over the course of the lifetimes of the individual partner lifetimes.

Yet again, it is difficult to pick out what counts as a case of a new or different holobiont without explicitly privileging one of the partners. I take this as a likely reason for why the concept of the holobiont rests on macrobe bias. Microbes are small. Macrobes are big. Microbes go where their macrobial associates go. Holobionts *seem* contiguous to us, in ways that symbiotic associations between macrobe-macrobe symbiotic associations like plants and pollinators don't. But this alone does not indicate that they are part of some larger whole. It is

just an artifact of their size.

Indexing holobiont identity to the host is not without its benefits. Focusing on the larger and longer lived host makes it easier to demarcate holobionts, a necessity for tracking holobiont changes over time and in response to environmental changes. Indexing symbiont community identity to the host is appropriate when the host is of primary interest[41]. But in the context of evolutionary individuality, all partners have equal weight, regardless of their size or longevity. Evolutionary individuals are only picked out by being entities that natural selection works on, not any physical or taxonomic features. Mistaking holobionts for units of selection appears to stem, in part, from host-centric thinking, macrobe bias and a reification of operational concepts like hologenome, microbiome, and metagenome[42].

4.6 Are Holobionts Organisms?

So far I have only discussed accounts of biological individuality that are tied to reproduction at the level of the holobiont. Another approach that has been proposed is based on David Hull's interactor account of individuality (Hull 1980, 1992), itself based on Richard Dawkins' replicator theory (Dawkins 1976). According to Hull, an interactor is "an entity that directly interacts as a cohesive whole with its environment in such a way that replication is differential" (Hull 1980). The replicators are entities which "pass on their structures largely intact from generation to generation" (Hull 1980). Replicators were originally conceived as an abstraction of the role of gene, while interactors are an abstraction of the role of organisms. The interactor is often identified as the primary unit of selection, but this role has also been extended

[41] See (Sterelny 2006) for a similar argument regarding the individuation of ecosystems.
[42] See Huss (2014) for an extended discussion and warning about reifying categories such as metagenome, microbiome and enterotype.

to the replicator (Lloyd 2012). The replicator-interactor framework has since been put to powerful use in the analysis of the complexities of inheritance and interaction in symbiotic consortiums (Sterelny 2001, 2004, 2011).

Criticism of the replicator-interactor framework has primarily focused on problems with replicator transmission and the reproduction of interactors (see for example, Griesemer 2000; Godfrey-Smith 2009). This has led some to suggest a notion of biological individuality that maintains interactors while decoupling them from replicators or particular reproductive requirements (O'Malley and Dupré 2009; Dupré 2012; Ereshefsky and Pedroso 2013, 2015). For example, Dupré (2012) makes the following claim: "…the organisms that are parts of evolutionary lineages are not the same things as the organisms that interact functionally with their biological and non-biological surroundings." This is immediately followed by a much stronger claim: "The latter, which I take to be more fundamental, are composed of a variety of the former, which are the more traditionally conceived organisms" (Dupré 2012).

The move to an interactor-only, or "updated interactor", concept of individuality is meant to capture the fact that many multilineage symbiotic consortiums appear to function as organisms in their environments, while also recognizing that the many replicators coming together within these consortiums are not inextricably tied together. It is suggested that all that is needed for an entity to be an interactor is enough interaction between the member parts such that the success or failure of the interactor has a unitary effect on the success or failure of its members (Ereshefsky and Pedroso 2013). For example, a higher survivorship in the members then if the members were living independently from each other (Ereshefsky and Pedroso 2013).

In the case of holobionts, much more needs to be said about the necessary type and

strength of interactions between holobiont partners such that the holobiont interacts as a functional whole (organisms) with the environment[43]. I will examine two approaches to organismality that may be compatible with the view that holobionts are modified interactors and biological individuals. The first is an approach that focuses on functional integration through metabolic collaboration. The second is the cooperation and conflict framework outlined in Queller and Strassmann (2009).

4.6.1 Functional Integration and Metabolic Dependency

Metabolic dependencies are a hallmark of the close symbiotic relationships that inspired the adoption of the holobiont framework. A type of biological individuality that emerges from lineage-forming entities collaborating in metabolism is a position explored by Dupré (2012) and Dupré and O'Malley (2009).

> "My colleague Maureen O'Malley and I (Dupré and O'Malley 2009) have suggested that the most fundamental way to think of living things is as the intersection of lineages and metabolism. The point we are making is that, contrary to the assumption that is fundamental to the one genome, one organism idea, the biological entities that form reproducing and evolving lineages are not the same as the entities that function as wholes in wider biological contexts. Functional biological wholes, the entities that we primarily think of as organisms, are in fact cooperating assemblies of a wide variety of lineage-forming entities." (Dupré 2012)

In their view of the natural world, "collaboration" among entities of fundamentally different types is essential to all living systems (Dupré and O'Malley 2009)[44]. Collaboration encompasses cooperation and competition, includes metabolic, structural, and fitness-affecting

---

[43] One reviewer suggested that many of the claims about holobionts as units of selection found in the quotations presented in the section entitled "Controversy about the Status of Holobionts" be interpreted as claims about holobionts being interactors. I disagree that this is the correct interpretation of the presented views. At the very least, it is unclear exactly what the quoted authors mean when they say holobionts are a unit of selection.

[44] Bouchard (2009) presents a similar view, arguing that "superindividuality" can emerge in persistent, functionally-integrated, multispecies communities.

interactions, and involves entities at many levels of biological organization. Dupré and O'Malley do not discuss holobionts, and so it is unclear whether they would endorse the position that holobionts are organisms, but their collaboration criteria for organismality seems a promising option for those wishing to argue that multilineage systems like holobionts are biological individuals.

Individual holobionts will almost inevitably contain partnerships that vary across the full range of collaboration as described by Dupré and O'Malley (2009). There are at least two reasons to be cautious about such a permissive approach to collaboration, if that collaboration is to be the glue that binds lower level individuals into a higher level individual.

First, recurring interactions, even ones with reciprocal benefits, needn't indicate that there is functional integration or active collaboration. Members of a particular host species will inevitably share similar physiologies, microbial defense mechanisms, and biochemistries compared to other species. The fact that selective microbial communities with the same composition are always associated with the same hosts may be due to differences in community assembly rather than cooperative behavior or a shared evolutionary history. Similar communities are to be expected across common similar environments.

Second, the evolution of metabolic dependencies or codependencies among host-microbe communities does not mean that the community, in this case a holobiont, is functionally integrated into a whole. When a nutrient is routinely provided by an organism's environment, selection for biochemical efficiency can lead to loss of genes in the particular biochemical pathway (Morris et al. 2012). The Black Queen Hypothesis states that this process is able to occur whenever there are microbial communities where "leaky products" are produced by

different members that are routinely associated with each other (Morris et al. 2012; Douglas and Werren 2016). Such processes can lead to interdependent communities without requiring selection or functional integration at the holobiont level (Sachs and Hollowell 2012; Douglas and Werren 2016). Mushegian and Ebert (2016) give plausible examples that include protective symbioses based on secondary metabolic functions, such as detoxification of heavy metals or plant toxins, or production of defensive compounds against other microbes, which are likely to be beneficial regardless of whether the microbe is in a host or non-host environment. Perhaps a more compelling example of the independence of metabolic dependency and functional integration is the mammalian gut. Mammalian digestive tracks provide microorganisms with all sorts of beneficial "leaky products" such as warmth, moisture and nutrients. Commensalist bacteria utilize and may even depend on those resources, while the host is entirely unaffected by the bacterial presence, by definition.

Is recurrence with metabolic integration enough to infer that holobionts are whole entities in their own right rather than mere associations of individuals or ephemeral communities? No, because ecological communities can have reproducible dynamics and predictable outcomes for their members without being the result of selection, integration or coevolution at the level of the community (Mushegian and Ebert 2016). Something further is needed to bind individuals into a whole. As argued in earlier sections, reproduction and cotransmission at the level of the whole are plausible options. Another possibility is cooperation.

4.6.2 Cooperation and Conflict

The second conception of organismality I will consider is the framework developed in

Queller and Strassmann (2009, see also Queller and Strassmann, this special issue[45]), which takes a social behavior approach to defining organisms. This approach is probably the most amendable to treating multi-species symbiotic consortiums like holobionts as organisms. Queller and Strassmann (2009) define organisms as "the largest unit of near-unanimous design." They fill this out by saying that "the organism is simply a unit with high cooperation and very low conflict among its parts. That is, the organism has adaptations and it is not much disrupted by adaptations at lower levels" (Queller and Strassmann 2009).

Conflicts of interest between the symbiont partners is a major obstacle to holobiont individuality. Within a single species, conflicts of interest can be suppressed by maintaining genetic homogeneity in the case of an individual multicellular organism, or high genetic relatedness in the case of a group of cooperating organisms. These routes are not available to multi-species holobionts.

Three factors that are important for conflict suppression in symbioses are: vertical transmission, specificity of symbiont relationships, and little or no dependency on a free-living state (Herre et al. 1999; Sachs et al. 2011; Lesser et al. 2013). Vertical transmission favors increased metabolic integration and can lead to symbiont genome reduction and obligate dependencies between partners, which in turn lowers conflict because the symbionts share a common fate (Sachs et al. 2011; Lesser et al. 2013). A high specificity in symbiont relationships reduces the number of competitive phenotypes that a host has to contend with, and reduces the conflict between individual microbes associated with the host. Little or no dependency on a free-

[45] Queller and Strassman (2016) argue that it is extremely unlikely that any holobionts qualify as organisms.

living state reduces the chance that a symbiont will have to contend with selective pressures from two environments, leading to specialized adaptation to symbiotic environments, and increasing the likelihood of coevolution and dependencies between partners.

These three factors can lead to the symbionts sharing common interests and an alignment of fitness. The difference between having mutually beneficial relationships and sharing common interests is a key difference between being just a group of interacting individuals and being a higher level individual. The open question is: how often does this happen?

We see again that there is a problem when we look at the holobiont as a whole unit. Some of the host-microbe relations might meet criteria for multi-species organismality, but it seems highly unlikely that *all* of the host-microbe relations will be cooperative. Even if there is cooperation without conflict between a host and all of its microbial symbionts, there will still inevitably be conflict in the holobiont between the microorganisms. For example, microbes will compete for resources within the host, including nutrients and space, as well as for access to the next host generation. The host will remain the site of a whole ecosystem of complex microorganismal interactions (see Mushegian and Ebert 2016 for a similar view). The idea that there is a single interaction between the host and its microbiome is an artifact of macrobe bias and the difficulty of gathering information about all the various host-microbe interactions. Epistemic limitations shouldn't tempt us to overly simplistic conclusions about biological individuality.

4.7 The Disunity of the Holobiont

It is unlikely that there is any holobiont that is also an evolutionary individual or organism if the holobiont is defined as a macrobe host and *all* of its associated microorganisms.

It is not impossible that a host and its symbionts could form a unit of selection, it is just that the conditions are unlikely to obtain. High partner fidelity and alignment of fitness are necessary. This is achieved by vertical inheritance or by strong mutual partner choice. Such high-fidelity associations are unlikely to occur across all of the partnerships within a holobiont. Where it does not, selective pressures at the level of the individual lineages will tend to put the partners into direct competition or active exploitation. Focusing on the processes, interactions, and relations that occur between holobiont partners like this opens up a suite of questions. Does vertical transmission lead to increased metabolic integration and alignment of fitness? Or are these necessary before vertical transmission becomes permanent? To what degree are holobiont partnerships species-specific coevolved consortiums vs. generalist assemblages taking advantage of leaky products or stable environments? What is the relationship between different biological parameters: mode of transmission vs. alignment of fitness vs. metabolic integration?

Many of these questions concern ecological relationships. As such, holobiont theory and research will be impoverished if it doesn't incorporate the powerful theoretical tools of community and ecosystem ecology. As we saw with the coral holobiont, holobionts are complex systems comprised of an array of lineages interacting in diverse ways. Holobionts are disunified in the sense that they share features of both individuals and communities. Some partner interactions are best considered as symbioses—ranging from mutualism to parasitism—where the partners mutually form a part of each other's environments. Other interactions long ago bound the individual lineages together into a higher-level lineage and evolutionary individual. I expect there are plenty of indeterminable cases on the road between ecological interaction and becoming a full-fledged individual. Holobionts are interesting because they share features of organisms and communities. Neither reducing the holobiont to a set of pairwise interactions

between symbiont partners nor treating the entire community as a single biological individual is a universally appropriate approach.

CONCLUSION

I conclude this dissertation with a summary of my main goals, and a look forward to possibilities for further developments. First, I hope I have clearly presented the evidence for aspects of biological complexity that demand an expanded account of scientific explanation. I explored how the complex dynamics, stochasticity, and varying degrees of orderliness of biological processes challenge views that all biological processes can be understood as mechanisms. I argued that explanation has two sides and paying attention to system-level features, like robustness, is a necessary complement to decompositional strategies that examine the workings of the system parts. I examined fields—ecology, cell, and developmental biology—where this two-part strategy has been independently developed. I also constructed a taxonomy of robustness that can be used across fields to shed light on how stability in biological systems is maintained.

When considering complex phenomena, or phenomena at scales vastly different than our own, it is important to keep in mind how our intuitions might lead us astray. Our causal reasoning habits have been formed in response to a world experienced at a medium scale. I examined how the world operates at much smaller scales, and how this gives us reason to be mindful of our casual reasoning habits, and doubtful about the applicability of our causal intuitions to the workings of living things at the microscopic and nano scales.

Finally, I took on an emerging debate within biology about the importance of microorganismal associations and the evolution of holobionts. It has been claimed that we need further change our understanding of the biological world through an update of the theory of evolution by natural selection. I showed how—contrary to the claims of many others—

holobionts are not evolutionary individuals. The ubiquity and strength of microorganismal associations with multi-cellular organisms requires that we reevaluate our picture of the biological world. But it doesn't change our theoretical commitments to how evolution by natural selection operates. Instead we should come to appreciate ourselves as walking ecosystems and communities of organisms.

The main ideas of this dissertation are framed around the idea that philosophical accounts of explanation need to be reevaluated and expanded in order to adequately capture the complexity of living systems and how they are investigated. The primary task of the first three chapters was to contribute to this ongoing project of building a better account of scientific explanation. The final chapter focused on claims that our understanding of biological entities, and possibly parts of possibly biological theory itself, need to be updated in response to recent empirical discoveries about the microbial world. Here at the close, I broaden my view and offer a few thoughts on the extent to which I think biology itself is also due for a reevaluation. A broad-scale reevaluation of the status and interrelations of biological theories has plausibly happened one time since Darwin introduced the theory of evolution by natural selection; it ended in the Modern Synthesis. There is considerable debate about whether or not this needs occur again in the form of an Extended Synthesis (Pigliucci 2007, 2009; Pigliucci and Muller 2010; Booth et al. 2016). It is argued that development, non-genetic inheritance, macro-evolutionary processes, ecology, and the role of contingency need to be better integrated into the theoretical structure set out by the Modern Synthesis (Pigliucci 2009). To this list could be added the role of microbial symbioses, lateral gene transfer, and the convergence of lineages.

Does biology need a new conceptual synthesis in order to integrate the current range of biological disciplines? I am sceptical that it does. Not because I don't think that previously

independent biological disciplines shouldn't be better integrated, but because I am sceptical that trying to subsume it all under a general unified theory will advance our knowledge of biological processes. I have repeated throughout my dissertation that I see explanatory strategies as heuristics, a set of tools that can be used to understand the world. I believe that biological theories are much same. Biology is pluralistic, with a wide range of methods and theoretical tools at its disposal. Whether all of these strategies can be unified under one general framework remains to be seen. Much of the natural world yields to our investigation as it is, without reference to a more general all-encompassing theory. Striving for a unified theory works as a good regulatory ideal, but we don't need that meta-theory to continue to make progress using a more loosely connected set of explanations, practices, methods, concepts, models, and theories.

REFERENCES

Achinstein P. 1983. The nature of explanation. Oxford (England): Oxford University Press.

Andersen H. 2012. The case for regularity in mechanistic causal explanation. Synthese. 189:415-432.

Apprill A, Marlow HQ, Martindale MQ, Rappe MS. 2009. The onset of microbial associations in the coral *Pocillopora meandrina*. ISME J. 3:685–699.

Aristotle. 1984. Categories. In: Barnes J, editor. The complete works of Aristotle, vol. 1. Princeton (NJ): Princeton University Press.

Astumian RD, Hänggi, P. 2002. Brownian motors. Physics Today. 55(11):33-39.

Astumian RD, Mukherjee S, Warshel A. 2016. The physics and physical chemistry of molecular machines. Chemphyschem 17(12): 1719-1741.

Austin JL. 1962. Sense and sensibilia. Oxford (England): Oxford University Press.

Bateson P, Gluckman P. 2012. Plasticity, robustness, development and evolution. New York (NY): Cambridge University Press.

Batterman RW. 2002. The devil in the details: asymptotic reasoning in explanation, reduction, and emergence. New York (NY): Oxford University Press.

Batterman RW. 2013. The tyranny of scales. In Batterman RW, editor. The Oxford handbook of philosophy of physics. New York (NY): Oxford University Press. p. 255-286.

Bechtel W. 2011. Mechanism and biological explanation. Philosophy of Science. 78(4):533-557.

Bechtel W. 2012. Understanding endogenously active mechanisms: a scientific and philosophical challenge. European Journal for Philosophy of Science. 2(2):233-248.

Bechtel W, Abrahamsen A. 2005. Explanation: a mechanist alternative. In: Studies in the History and Philosophy of the Biological and Biomedical Sciences. 36:421-441.

Bechtel W, Abrahamsen A. 2013. Thinking dynamically about biological mechanisms: networks of coupled oscillators. Foundations of Science. 18:707-723.

Bechtel W, Richardson RC. 1993/2010. Discovering complexity: decomposition and localization as strategies in scientific research. Reprint. Cambridge (MA): MIT Press.

Berg RD. 1996. The indigenous gastrointestinal microflora. Trends Microbiology 4:430–435.

Berridge MJ, Bootman MD, Roderick HL. 2003. Calcium signalling: dynamics, homeostasis and remodelling. Nature Reviews: Molecular Cell Biology. 4:517-529.

Bocci V. 1992. The neglected organ: bacterial flora has a crucial immunostimulatory role. Perspect Biol

Med. 35(2):251-60.

Booth A. 2014. Symbiosis, selection and individuality. Biology & Philosophy. 29:657-673.

Booth A, Mariscal C, Doolittle WF. 2016. The modern synthesis in the light of microbial genomics. The Annual Review of Microbiology. 70:279-297.

Bordenstein SR, Theis KR. 2015. Host biology in light of the microbiome: ten principles of holobionts and hologenomes. PLoS Biology. 13(8):e1002226.

Bosch TCG, McFall-Ngai MJ. 2011. Metaorganisms as the new frontier. Zoology. 114:185-190.

Bouchard F. 2013. What is a symbiotic superindividual and how do you measure its fitness? In Bouchard F and Huneman P, editors. From groups to individuals: evolution and emerging individuality. Cambridge (MA): MIT Press. p. 243-264.

Boudry M, Pigliucci M. 2013. The mismeasure of machine: synthetic biology and the trouble with engineering metaphors. Studies in History and Philosophy of Biological and Biomedical Sciences. 44(4):660-668.

Braun E, Marom S. 2015. Universality, complexity and the praxis of biology: two case studies. Studies in History and Philosophy of Biological and Biomedical Sciences. 53:68-72.

Bright M, Bulgheresi S. 2010. A complex journey: transmission of microbial symbionts. Nature Reviews: Microbiology. 8(3):218-230.

Bromberger S. 1961. The concept of explanation [dissertation]. [Cambridge (MA)]: Harvard University.

Bromberger S. 1965. An approach to explanation. In: Butler RJ, editor. Analytical philosophy, volume 2, p. 72–105. Oxford (England): Oxford University Press.

Brucker RM, Bordenstein SR. 2014. Response to comment on “The hologenomic basis of speciation: gut bacteria cause hybrid lethality in the genus Nasonia.” Science 345:1011.

Buss L. 1987. The evolution of individuality. Princeton (NJ): Princeton University Press.

Clarke E. 2011. The problem of biological individuality. Biological theory. 5(4):312-325.

Clarke E. 2016. Levels of selection in biofilms: multispecies biofilms are not evolutionary individuals. Biology & Philosophy. 31:191-212.

Craver CF. 2007. Explaining the brain: mechanisms and the mosaic unity of neuroscience. Oxford (England): Clarendon Press.

Dashtia A, Schwandera P, Langloisb R, Funga R, Lib W, Hosseinizadeha A, Liaob HY, Pallesenc J, Sharmab G, Stupinad VA, et al. 2014. Trajectories of the ribosome as a brownian nanomachine, PNAS. 111(49):17492–17497.

Dawkins R. 1976. The Selfish Gene. New York (NY): Oxford University Press.

Douglas AE. 2008. Conflict, cheats and the persistence of symbioses. New Phytologist. 177:849–58.

Douglas AE, Werren JH. 2016. Holes in the hologenome: why host-microbe symbioses are not holobionts. mBio. 7(2):e02099-15.

Drake S. 1981. Cause, experiment and science. Chicago (IL): University of Chicago Pres.

Dupré J. 2012. The polygenomic organism. In: Dupré J, editor. Processes of life: essays in the philosophy of biology. Oxford (England): Oxford University Press. p. 116–127.

Dupré J, O'Malley MA. (2009. Varieties of living things: life at the intersection of lineage and metabolism. Philosophy and Theory in Biology. 1.

Ereshefsky M, Pedroso M. 2013. Biological individuality: the case of biofilms. Biology & Philosophy. 28(2):331–349.

Ereshefsky M, Pedroso M. 2015. Rethinking evolutionary individuality. PNAS 112:10126-10132.

Feldhaar H. 2011. Bacterial symbionts as mediators of ecologically important traits of insect hosts. Ecological Entomology. 36:533-543.

Garcia JR, Gerardo NM. 2014. The symbiont side of symbiosis: do microbes really benefit? Frontiers in Microbiology. 5(510):1-6.

Gilbert SF. 2014. Symbiosis as the way of eukaryotic life: the dependent co-origination of the body. J Biosci 39:201–209.

Gilbert SF, Sapp J, Tauber AI. 2012. A symbiotic view of life: we have never been individuals. Quarterly Review of Biology. 87(4):325-341.

Glennan S. 1996. Mechanisms and the nature of causation. Erkenntnis. 44:50-71.

Glennan S. 1997. Probable causes and the distinction between subjective and objective chance. Noûs. 31(4):496-519.

Glennan S. 2002. Rethinking mechanistic explanation. Philosophy of Science. 69:S342-S353.

Godfrey-Smith P. 2009. Darwinian populations and natural selection. New York (NY): Oxford University Press.

Godfrey-Smith P. 2011. Agents and acacias: replies to Dennett, Sterelny, and Queller. Biology & Philosophy. 26:501.

Godfrey-Smith P. 2013. Darwinian individuals. In: Bouchard F, Huneman P, editors. From groups to individuals: evolution and emerging individuality. Cambridge (MA): MIT Press. p. 17–36.

Godfrey-Smith P. 2015. Reproduction, symbiosis, and the eukaryotic cell. PNAS. 112(33):10120-10125.

Godfrey-Smith P. 2016. Mind, matter, and metabolism. Journal of Philosophy.

Golubic S, Radtke G, Le Campion-Alsumard T. 2005. Endolithic fungi in marine ecosystems. Trends Microbiology. 13:229–235.

Gordon J, Knowlton N, Relman DA, Rohwer F, Youle M. 2013. Superorganisms and holobionts. Microbe. 8(4):152-153.

Green S. 2015. Can biological complexity be reverse engineered? Studies in History and Philosophy of Biological and Biomedical Sciences. 53: 73-83.

Green S, Jones N. 2016. Constraint-based reasoning for search and explanation: strategies for understanding variation and patterns in biology. Dialectica. 70(3):343-374.

Griesemer J. 2000. Development, culture and the units of inheritance. Philosophy of Science. 67:S348–S368.

Griesemer J. 2014. Reproduction and scaffolded developmental processes: an integrated evolutionary perspective. In: Minelli A, Pradeu T, editors. Towards a theory of development. New York (NY): Oxford University Press. p. 183-202.

Griesemer J. 2016. Reproduction in complex life cycles: toward a developmental reaction norms perspective. Philosophy of Science. online first.

Guay A, Pradeu T. 2016. Progressive steps towards a unified conception of individuality across the sciences. In: Guay A, Pradeu T, editors. Individuals across the sciences. New York (NY): Oxford University Press.

Hall N. 2004. Two concepts of causation. In: Collins J, Hall N, Paul LA, editors. Causation and counterfactuals. Cambridge (MA): MIT Press. p. 225-76.

Hattab H. 2011. The mechanical philosophy. In: Clarke DM, Wilson C, editors. The Oxford handbook of philosophy in early modern Europe. Oxford (England): Oxford University Press.

Hempel CG. 1965. Aspects of scientific explanation. In: Hempel CG, editor. Aspects of scientific explanation and other essays in the philosophy of science. New York (NY): Macmillan. p. 331–496.

Herre EA, Knowlton N, Mueller UG, Rehner SA. 1999. The evolution of mutualisms: exploring the paths between conflict and cooperation. Trends in Ecology Evolution. 14:49–53.

Hoffman P. 2012. Life's Ratchet. New York (NY): Basic Books.

Holling CS. 1973. Resilience and stability of ecological systems. Annual Review of Ecology and Systematics. 4:1-23.

Howick J. 2011. Exposing the vanities—and a qualified defense—of mechanistic reasoning in health care decision making. Philosophy of Science. 78(5):926-940.

Hull D. 1978. A matter of individuality. Philosophy of Science. 45(3):335-360.

Hull D. 1980. Individuality and selection. Annual Review of Ecology and Systematics. 11:311-332.

Hull D. 1992. Individual. In: Fox-Keller E, Lloyd E, editors. Keywords in evolutionary biology. Cambridge (MA): Harvard University Press. p. 181-187.

Hume BCC, Voolstra CR, Arif C, D'Angelo C, Burt JA, Eyal G, Loya Y, Wiedenmann J. 2016. Ancestral genetic diversity associated with the rapid spread of stress-tolerant coral symbionts in response to Holocene climate change. PNAS. 113(16):4416-4421.

Hume D. 1748. An enquiry concerning human understanding. London (England): A Millar.

Huss J. 2014. Methodology and ontology in microbiome research. Biol Theory. 9:392-400.

Kitano H. 2004. Biological robustness. Nature Reviews Genetics. 5(11):826-837.

Kitano H. 2007. Towards a theory of biological robustness. Molecular Systems Biology. 3:137.

Kitcher P. 1981. Explanatory unification. Philosophy of Science. 48:507–31.

Knowlton N, Rohwer F. 2003. Multispecies microbial mutualisms on coral reefs: the host as a habitat. The American Naturalist. 162:S51-S62.

Lesser MP, Stat M, Gates RD. 2013. The endosymbiotic dinoflagellates (*Symbiodinium* sp.) of corals are parasites and mutualists. Coral Reefs. 32:603-611.

Levin SA, Lubchenco J. 2008. Resilience, robustness, and marine ecosystem-based management. BioScience. 58(1):27-32.

Levy A. 2013. Three kinds of new mechanism. Biology & Philosophy. 28:99-114.

Levy A. 2014. Machine-likeness and explanation by decomposition. Philosopher's Imprint. 14(6).

Levy A, Bechtel W. 2013. Abstraction and the organization of mechanisms. Philosophy of Science. 80(2):241-261.

Lewis D. 1986. Causal explanation. Philosophical Papers. 2:214-40.

Lewontin RC. 1970. The units of selection. Annual Review of Ecology and Systematics. 1(1):1–18.

Lewontin RC. 1974. The genetic basis of evolutionary change. New York (NY): Columbia University Press.

Lewontin RC. 1978. Adaptation. Scientific American. 239(3):212-229.

Lloyd E. 2012. Units and levels of selection. In: Zalta EN, editors. The Stanford Encyclopedia of Philosophy. Winter 2012. http://plato.stanford.edu/archives/win2012/entries/selection-units/

Machamer P, Darden L, Craver, CF. 2000. Thinking about mechanisms. Philosophy of Science. 67:1-25.

Margulis L. 1991. Symbiogenesis and symbionticism. In: Margulis L, Fester R, editors. Symbiosis as a source of evolutionary innovation: speciation and morphogenesis. Cambridge (MA): MIT Press. p. 1-14.

Maynard Smith J, Szathmary E. 1995. The major transitions in evolution. Oxford (England): Oxford University Press.

McFall-Ngai M. 2016. Introduction to the hologenome special series. mBio. 7(2):e00371-16.

McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Loso T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, et al. 2013. Animals in a bacterial world, a new imperative for the life sciences. PNAS. 110:3229–3236.

Michod RE. 1999. Darwinian dynamics. Evolutionary transitions in fitness and individuality. Princeton (NJ): Princeton University Press.

Mill JS. 1882. A system of logic, ratiocinative and inductive: being a connected view of the principles of evidence and the methods of scientific investigation. London (England): Longmans, Green, and Co.

Moore PB. 2012. How should we think about the ribosome? Annual Review of Biophysics. 41:1-19.

Moran NA. 2006. Symbiosis. Current Biology. 16(October):R866–R871.

Moran NA, Sloan DB. 2015. The hologenome concept: helpful or hollow? PLoS Biology 13(12):e1002311.

Moreno A, Ruiz-Mirazo K, Barandiaran X. 2009. The impact of the paradigm of complexity on the foundational frameworks of biology and cognitive science. In: Hooker C, editor. Handbook of the philosophy of science, volume 10: philosophy of complex systems. General editors: Gabbay DM, Thagard P, Woods J.

Moreno A, Mossio M. 2015. Biological autonomy: A philosophical and theoretical enquiry. New York (NY): Springer.

Morris JJ, Lenski RE, Zinser ER. 2012. The black queen hypothesis: evolution of dependencies through adaptive gene loss. mBio. 3(2):e00036-12.

Mossio M, Bich L. 2014. What makes biological organization teleological? Synthese.

Mossio M, Bich L, Moreno A. 2013. Emergence, closure and inter-level causation in biological systems. Erkenntnis. 78(Supplement 2): 153.

Mushegian AA, Ebert D. 2016. Rethinking “mutualism” in diverse host-symbiont communities. Bioessays. 38:100-108.

Nathan MJ. 2014. Causation by concentration. British Journal for the Philosophy of Science 55(2):191-212.

Nicholson DJ. 2012. The concept of mechanism in biology. Studies in History and Philosophy of Biological and Biomedical Sciences. 43:152-163.

Nijhout HF. 2002. The nature of robustness in development. BioEssays. 24:553-563.

O'Malley MA. 2016. Reproduction expanded: multigenerational and multilineal units of evolution. Philosophy of Science. 83(5): 835-847.

Pigliucci M. 2007. Do we need an extended evolutionary synthesis? Evolution. 61:2743–2749.

Pigliucci M. 2009. An extended synthesis for evolutionary biology. Ann. N.Y. Acad. Sci. 1168:218–228.

Pigliucci M, Boudry M. 2011. Why machine-information metaphors are bad for science and science education. Science & Education. 20:453-471.

Pigliucci M, Muller G, editors. 2010. Evolution: the extended synthesis. Cambridge (MA): MIT Press.

Pradeu T. 2010. What is an organism? An immunological answer. History and Philosophy of the Life Sciences. 32(2–3):247–267.

Pradeu T. 2012. The limits of the self: immunology and biological identity. Vitanza E, translator. New York (NY): Oxford University Press.

Queller DC, Strassmann JE. 2009. Beyond society: the evolution of organismality. Phil Trans R Soc B 364:3143-3155.

Queller DC, Strassmann JE. 2016. Problems of multi-species organisms: endosymbionts to holobionts. Biology & Philosophy. 31:855.

Railton P. 1981. Probability, explanation, and information. Synthese. 48:233-56.

Rohwer F, Seguritan V, Azam F, Knowlton N. 2002. Diversity and distribution of coral-associated bacteria. Marine Ecology Progress Series. 243:1-10.

Rosenberg E, Koren O, Reshef L, Efrony R, Zilber-Rosenberg I. 2007. The role of microorganisms in coral health, disease and evolution. Nature Reviews Microbiology. 5:355–362.

Roth MS. 2014. The engine of the reef: photobiology of the coral-algal symbiosis. Frontiers in Microbiology. 5(422):1-22.

Sachs JL, Essenberg CJ, Turcotte MM. 2011. New paradigms for the evolution of beneficial infections. Trends Ecol Evol. 26:202–209.

Sachs JL, Hollowell AC. 2012. The origins of cooperative bacterial communities. mBio 3(3):e00099-12

Salmon WC. 1984. Scientific explanation and the causal structure of the world. Princeton (NJ): Princeton University Press.

Salmon WC. 1970. Statistical explanation. In Colodny RG, editor. The nature and function of scientific theories. Pittsburgh (PA): University of Pittsburgh Press.

Salmon WC. 1971. Statistical explanation and statistical relevance. Pittsburgh (PA): University of Pittsburgh Press.

Schmaltz TM. 2011. From causes to laws. In: Clarke DM, Wilson C, editors. The Oxford handbook of philosophy in early modern Europe. Oxford (England): Oxford University Press.

Simon H. 1962. The architecture of complexity. PNAS. 106(6):467-482.

Singh Y, Ahmad J, Musarrat J, Ehtesham NZ, Hasnain SE. 2013. Emerging importance of holobionts in evolution and in probiotics. Gut Pathology. 5:12.

Skillings D. 2015. Mechanistic explanation of biological processes. Philosophy of Science. 82:1-13.

Skillings D. 2016. Holobionts and the ecology of organisms: Multi-species communities or integrated individuals? Biology & Philosophy. 31(6):875-892.

Smith JEH. 2011. Machines, souls, and vital principles. In: Clarke DM, Wilson C, editors. The Oxford handbook of philosophy in early modern Europe. Oxford (England): Oxford University Press.

Sober E, Wilson DS. 1998. Unto others: the evolution and psychology of unselfish behaviour. Cambridge (MA): Harvard University Press.

Stat M, Baker AC, Bourne DG, Correa AMS, Forsman Z, Huggett MJ, Pochon X, Skillings D, Toonen RJ, van Oppen MJH, Gates RD. 2012. Molecular delineation of species in the coral holobiont. In: Lesser M, editor. Advances in Marine Biology, Vol. 63. Amsterdam (The Netherlands): Academic Press. p. 1-65.

Steinacher A, Soyer OS. 2012. Evolutionary principles underlying structure and response dynamics of cellular networks. In: Soyer OS, editor. Evolutionary Systems Biology. New York (NY): Springer.

Sterelny K. 1996. Explanatory pluralism in evolutionary biology. Biology & Philosophy. 11:193-214.

Sterelny K. 2001. Niche construction, developmental systems, and the extended replicator. In: Oyama S, Griffiths PE, Gray RD, editors. Cycles of contingency: developmental systems and evolution. Cambridge (MA): MIT Press. p. 333-349.

Sterelny K. 2004. Symbiosis, evolvability and modularity. In: Schlosser G, Wagner G, editors. Modularity in development and evolution. Chicago (IL): University of Chicago Press.

Sterelny K. 2006. Local ecological communities. Philosophy of Science. 73:215-231.

Sterelny K. 2011. Darwinian spaces: Peter Godfrey-Smith on selection and evolution. Biology & Philosophy. 26:489–500.

Strevens M. 2008. Depth: an account of scientific explanation. Cambridge (MA): Harvard University

Press.

Strevens M. 2012. Précis of depth. Philosophy and Phenom Research. 84(2):447-460.

Theis KR, Dhelly NM, Klassen JL, Brucker RM, Baines JF, Bosch TCG, Cryan JF, Gilbert SF, Goodnight CJ, Lloyd EA, et al. 2016. Getting the hologenome concept right: an eco-evolutionary framework for hosts and their microbiomes. mSystems. 1(2):e00028-16.

Thompson JR, Rivera HE, Closek CJ, Medina M. 2015. Microbes in the coral holobiont: partners through evolution, development, and ecological interactions. Frontiers in Cellular and Infection Microbiology. 4(176):1-20.

van Fraassen BC. 1980. The scientific image. New York (NY): Oxford University Press.

Wagner A. 2005. Distributed robustness versus redundancy as causes of mutational robustness. BioEssays. 27:176-188.

Wegley L, Edwards R, Rodriguez-Brito B, Liu H, Rohwer F. 2007. Metagenomic analysis of the microbial community associated with the coral *Porites astreoides*. Environmental Microbiology. 9(11):2707-2719.

Whitacre JM. 2012. Biological robustness: paradigms, mechanisms, and systems principles. Frontiers in Genetics. 3(67):1-15.

Wimsatt WC. 1972. Complexity and organization. In: Schaffner KF, Cohen RS, editors. PSA 1972, Proceedings of the Philosophy of Science Association. Dordrecht (The Netherlands): Reidel. p. 67-86.

Wimsatt WC. 1981. Robustness, reliability and overdetermination. In: Brewer M, Collins B, editors. Scientific Inquiry and the Social Sciences. San Francisco (CA): Jossey-Bass. p. 124-163.

Wimsatt WC. 1986. Forms of aggregativity. In: Donagan A, Perovich AN, Wedin MV, editors. Human Nature and Natural Knowledge. Dordrecht (The Netherlands): Reidel. p. 259–93.

Wimsatt WC. 1994. Ontology of complex systems. Canadian Journal of Philosophy. 20:207-274.

Wimsatt WC. 2007. Re-engineering philosophy for limited beings: piecewise approximations to reality. Cambridge (MA): Harvard University Press.

Woodward J. 2003. Making things happen: a theory of causal explanation. New York (NY): Oxford University Press.

Woodward J. 2010. Causation in biology: stability, specificity, and the choice of levels of explanation. Biology & Philosophy. 25:287-318.

Woodward J. 2013. Mechanistic explanation: its scope and limits. Proceedings of the Aristotelian Society Supplementary. 87:39-65.

Woodward J. 2014. A defense of the legitimacy of causal thinking by reference to the only standard that matters – usefulness. PSA Presidential Address.

Wooldridge SA. 2010. Is the coral-algae symbiosis really ‘mutually beneficial’ for the partners? Bioessays. 32:615–625.

Xu J, Gordon JI. 2003. Honor thy symbionts. PNAS. 100:10452–10459.

Zilber-Rosenberg I, Rosenberg E. 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiology Reviews. 32:723–735.

The End.