



Københavns Universitet

Constraint-Based Reasoning for Search and Explanation

Green, Sara; Jones, Nicholas

Published in:
Dialectica

DOI:
[10.1111/1746-8361.12145](https://doi.org/10.1111/1746-8361.12145)

Publication date:
2016

Citation for published version (APA):
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. <https://doi.org/10.1111/1746-8361.12145>

Constraint-Based Reasoning for Search and Explanation: Strategies for Understanding Variation and Patterns in Biology

Sara Green

Department of Science Education, University of Copenhagen, Email: sara.green@ind.ku.dk

Nicholaos Jones

Department of Philosophy, University of Alabama in Huntsville, Huntsville AL 35899 USA
Email: nick.jones@uah.edu

Abstract

Life scientists increasingly rely upon abstraction-based modeling and reasoning strategies for understanding biological phenomena. We introduce the notion of constraint-based reasoning as a fruitful tool for conceptualizing some of these developments. One important role of mathematical abstractions is to impose formal constraints on a search space for possible hypotheses and thereby guide the search for plausible causal models. Formal constraints are, however, not only tools for biological explanations but can be explanatory by virtue of clarifying general dependency-relations and patterning between functions and structures. We describe such situations as *constraint-based explanations* and argue that these differ from mechanistic strategies in important respects. While mechanistic explanations emphasize change-relating causal features, constraint-based explanations emphasize formal dependencies and generic organizational features that are relatively independent of lower-level changes in causal details. Our distinction between mechanistic and constraint-based explanations is pragmatically motivated by the wish to understand scientific practice. We contend that delineating the affordances and assumptions of different explanatory questions and strategies helps to clarify tensions between diverging scientific practices and the innovative potentials in their combination. Moreover, we show how constraint-based explanation integrate several features shared by otherwise different philosophical accounts of abstract explanatory strategies in biology

Introductory Remarks

The life sciences increasingly rely upon abstraction-based modeling and reasoning strategies for managing and understanding biological complexity. New experimental techniques and theoretical insights encourage mathematical and computational aids for interpreting and organizing biological data. Meanwhile, explanatory ideals from engineering, systems theory, and physics influence new approaches – such as systems biology – that aim, in part, to identify and formalize so-called design principles. These principles are typically defined at a high level of abstraction, such that the same principle applies to engineered and biological systems alike.¹ Examples include feedback principles, bi-stable switching, as well as general network properties such as scale-free distribution and network motifs (Alon, 2007; Csete & Doyle, 2002).

How these strategies relate to mechanistic accounts of discovery and explanation is currently a hot topic in philosophy of biology. Some advocates of mechanistic accounts take mathematical models to be non-explanatory ‘how-possibly’ models for constructing spaces of possible mechanisms (Craver, 2007; Matthiessen, 2015). Others, recognizing the growing mathematical embedding of the life sciences, extend their conception of mechanistic explanation in order to accommodate the mathematical aspects of dynamical models (Bechtel & Abrahamsen, 2011, 2012; Brigandt, 2013; Fagan, 2012; Levy & Bechtel, 2013). In contrast to these approaches, an alternative is to describe some uses of mathematical abstractions as instantiations of an additional type of explanation in biology (Brailard, 2010; Gross, 2015; Huneman, 2010; Morange, 2011). This alternative approach emphasizes that formal principles in biology cite generic dependencies among various structural and functional factors holding for all members of a class of systems, regardless of the of causal details characterizing these systems. It maintains that general laws and principles – rather than causal, mechanistic details – carry the explanatory burden, even when the explananda of interest are causally mediated. There appears, accordingly, to be a conflict in the existing philosophical literature, between those who prioritize the explanatory force of mechanistic detail and those who emphasize the irrelevance of this same detail.

Our view is that these different approaches to the role of mathematical abstractions can co-exist if one is willing to admit a certain diversity of explanatory strategies in biology. Our goal is to articulate a foundation for the compatibility of these strategies. Rather than focusing on whether the explanations are causal or non-causal, we examine various uses of abstract, formal principles in biological practice. In particular, we aim to clarify the reasons for why some scientists prefer more abstract representational frameworks even if more detailed models are available or possible.² We present a set of case studies showing that formal mathematical principles play at least two complementary roles.

¹ The search for design principles needs not imply assumptions about a design process such as natural selection. For a discussion on this issue, see (Green, Levy, & Bechtel, 2015)

² We are not claiming that mathematical models are always developed with explanatory aims in mind. Other important purposes of mathematical modeling are to facilitate prediction and control for pragmatic purposes. We believe that formal constraints play an important role for these aims as well, but we do not develop that argument here.

The first resembles Craver's idea that abstract 'how-possibly' models facilitate developing more detailed and more accurate 'how-actually' models that explain how specific phenomena are causally produced. While acknowledging the merits of this description in many contexts, we argue that there are explanatory aims in biology other than explaining the causal processes of concrete systems. In particular, determining the space of possible models that can realize a particular function, or explaining why the same generic model applies to different contexts, are sometimes prioritized over explaining the workings of specific systems. In such situations, the focus shifts from strategies for discovering mechanistic details to strategies for determining whether classes of potential detail are within the realm of biological plausibility (or possibility) or of a specific *type*. This requires suppressing mechanistic detail and reasoning about the influence of constraints, often via a formalization of factors that simultaneously limit and enable certain functional characteristics of living systems. Thus, causal details are de-emphasized in order to afford an understanding of the conditions that delimit what is actually *possible* (see Section 3). In this second role, formal principles help to answer questions about *what* kind of biological "designs" or structures are possible relative to a well-defined set of constraints holding for all systems of a particular type, and *why*. We argue that answers to such questions should be seen as a distinct, but complementary, type of explanation. Because this is a controversial issue in philosophy of biology, showing the relevance of these explanations to biology will be the main focus of our paper. We argue that constraint-based reasoning has virtues that mechanistic explanations do not have and vice versa. In particular, the former can help biologists understand the constraints on biological variation and thereby to clarify why different systems share certain patterns and organizational features.

We proceed as follows. We first introduce the notion of constraint and introduce the different roles that constraint-based reasoning plays in biology. Section 2 discusses the heuristic role of abstraction and mathematical modeling in constraint-based search, and we illustrate this kind of heuristic strategy with a recent episode from biological practice. We return to the issue of explanatory salience in Section 3, discussing ways in which different explanatory strategies prioritize or deemphasize mechanistic detail. In Section 4, we provide a template for what we call constraint-based explanation. We use this template to highlight important features of recently formulated types of explanations that de-emphasize the significance or relevance of causal details related to the component parts of the system to be explained. Section 5 defends the autonomy of constraint-based explanations, and Section 6 offers concluding remarks.

1. Constraint-based reasoning

In the following we shall argue that the aforementioned explanatory diversity can be clarified through attention to the distinct roles of formal constraints. Constraints are conditions - or representations of conditions - that both limit and afford a certain scope of possible structures and functions that can be instantiated in a system of a particular type. Some constraints are physically imposed upon the structure and function of natural systems. For example, a skeleton's rigidity is a natural constraint that limits the flexible functionality of body parts but also affords upright movement on land. Other constraints are formal, imposed by us upon our models of natural systems, often with an aim of approximating the natural constraints present for those systems. For

example, a mathematical equation characterizing a skeleton as perfectly inflexible is a formal constraint that approximates the skeleton's less than perfect rigidity.

A formal constraint, as we understand the notion, is a general principle, often conceptualized in mathematical terms, that characterizes a dependency relation among some of a system's (natural or hypothesized) features in a way that limits and affords a certain set of possibilities for our models of the system. Some formal constraints are law-like generalizations, such as Fick's law of gas diffusion or geometrical surface-volume ratios. Whereas these generalizations are not specifically biological, they enable us to understand generalizable constraints on biological variation such as allometric scaling. For instance, body proportions, metabolic rates, or feasible respiratory strategies for a given animal are not only given by phylogenetic variance but also by constraints considering the size of the animal (Green, 2015; Wouters, 2007). We provide further examples below, and hope that the few mentioned here suffice in illustrating how formal constraints, as we understand them, differ from mechanistic details: they cite formal dependency relations; their referents (but not necessarily their relata), if such there be, are not *specific* biological systems; rather they often apply to many systems with diverse mechanistic details.

Reasoning with formal constraints affords insights to why certain functional or organizational patterns are common or absent in nature. Formal constraints are central to theorizing about what certain types of systems have in common and why. They also facilitate representing abstract features of our guiding hypotheses about systems, such as that a function is optimally or suboptimally performed (see Section 2). What we call constraint-based reasoning is akin to what Nersessian (2008, 193) calls *generic abstraction* that involves "selectively suppressing information instantiated in a representation, so as to make inferences that pertain only to the generic case" (Nersessian 2008, 193). Generic cases are abstractions specifying what all systems of a particular type have in common, e.g. organizational features that are shared by all harmonic oscillators, electromagnetic fields, biological networks etc. and that make some causal details irrelevant for understanding why the systems exhibit a general pattern. Our claim is that some explanatory strategies in the life sciences prioritize formal constraints and generic patterns as more explanatorily salient than mechanistic details. In doing so, these strategies produce a kind of explanation – what we shall call *constraint-based explanation* – that uses abstractions and mathematical models in a way that mechanistic explanations do not. We argue that this diversity of explanatory weighting does not lead to a conflict regarding the general importance or irrelevance of mechanistic details. Because there are different questions one can raise about biological systems, a complete understanding requires different types of explanation that we aim to clarify in this paper.

Our claim is not that *only* constraint-based explanations use formal constraints. Quite the contrary. Strategies that aim to provide (basic) mechanistic explanations sometimes engage in what we shall call *constraint-based search*, invoking formal constraints to guide the search for plausible mechanistic detail. Dynamical mechanistic explanations invoke the dynamic properties of computational models, and these models are explanatory insofar as they are grounded in relevant mechanistic details. Formal constraints, such as mass action rate laws or Michaelis-Menten kinetics, figure in these explanations as warrants for inferring mathematical models from relevant mechanistic detail (Jones & Wolkenhauer, 2012, 712-718). These constraints, however,

are only explanatory insofar as they can become closely integrated with causal details on properties of parts that make up their mechanistic underbelly (Bechtel 2011). Constraint-based explanations, as we shall argue, invert this explanatory weighting, producing explanations where delineation of the space of the possible is prior to ‘how-actually’ models for specific systems. We summarize these differences in Figure 1.

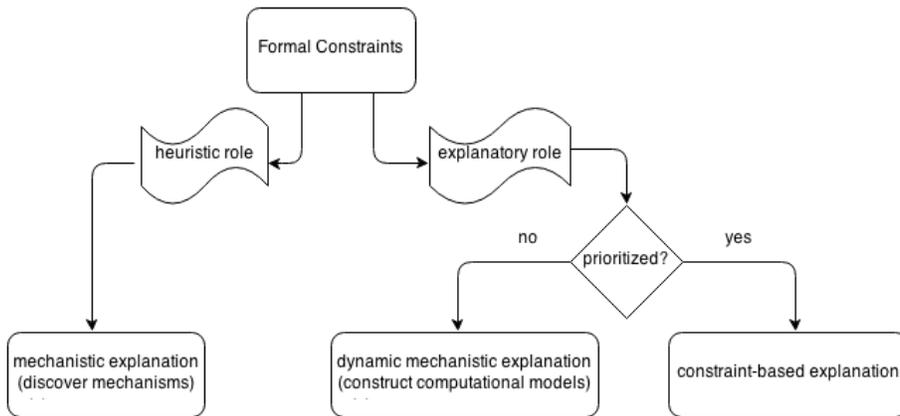


Figure 1: Some heuristic and explanatory roles of formal constraints in biology. Unlike traditional mechanistic accounts, dynamic mechanistic explanations and constraint-based explanations acknowledge an explanatory role of formal constraints. In constraint-based explanations, formal constraints are prioritized over mechanistic details.³

As illustrated in the figure, we distinguish between heuristic and explanatory roles of formal constraints. We do not believe that there is any sharp distinction to be drawn in scientific practice as scientists rarely have or need explicit criteria defining when an explanation has been reached. The intent of our distinction is to emphasize how formal constraints play distinct roles in biological research. The notion of research heuristics allows us to characterize formal constraints as akin to (and supportive of) other mechanistic heuristics used in the life sciences such as decomposition, localization, analogical reasoning (Bechtel and Richardson, 1993; Hesse, 1963). With the explanatory role of constraint-based reasoning we, however, wish to emphasize that constraint-based reasoning in biological research is not reducible to this role (cf. Matthiessen, 2015). These distinct roles will in the following be illustrated with concrete examples from biological research.

2. Constraint-Based Search for Possible Hypotheses and Models

Scientific problem-solving typically requires heuristic strategies for minimizing search within a vast space of possible solutions and thereby directing testing efforts to a narrower range of candidates (Simon, 1977). There are many such strategies for finding plausible mechanistic models, including analogical reasoning, decomposition and localization, and so on. Reasoning with formal constraints provides another such strategy. When available data is either too complex or too incomplete to feasibly direct resource-intensive testing efforts to a narrower range of

³ Constraints play heuristic, as well as explanatory roles, for developing dynamic mechanistic explanations and constraint-based explanations. For simplicity, we omit this aspect in the diagram.

candidates, formal constraints make search through the space of possibilities tractable. They do so by acting as adequacy conditions, excluding from further consideration those models that do not satisfy the relevant constraints. We illustrate this heuristic role of formal constraints with an example from recent research on robustness.

2.1. Robustness

Biological robustness, roughly characterized, is a capacity for a biological system to maintain its performance across a range of perturbations to internal and external conditions (Kitano, 2007). This capacity occurs across many scales, from genomes that are robust by virtue of repair systems that reduce replication errors to ecosystems that are robust by virtue of resisting the extinction of keystone species (Lenski, Barrick, & Ofria, 2006). Biological systems achieve robustness through several prominent mechanisms: feedback loops that maintain dynamic responses; alternative mechanisms that compensate for failure elsewhere through redundancy, or mechanistic modularity that isolates perturbative effects (Kitano, 2004, 827-828). Not every system function is robust to every perturbation: systems are sensitive to at least some variations of their inputs. But biologists are increasingly recognizing that robustness is ubiquitous (Kitano 2004, 829).

Biologists have a general expectation that many functions within biological systems will be robust to relatively wide ranges of perturbation, and an increasing number of researchers highlight the fruitfulness of framing biological robustness in a conceptual framework of mathematical rigor. Whence, Stelling and colleagues argue that “biological systems are robust—therefore an appropriate mathematical representation must also be robust” (Stelling et al. 2004). Stelling and colleagues infer, from their criterion requiring that appropriate representations mirror system robustness, that “mechanistic details that are necessary to satisfy this criterion form testable hypotheses” (Stelling et al. 2004). Their idea, in our terminology, is formalized criteria for robustness point to necessary or plausible features of a model or a system realizing this capacity.

In order to illustrate this idea, and the broader notion of constraint-based search, we provide a case study in which scientists appeal to a criterion of parametric robustness in order to support judgments about the relative plausibility of different mechanisms. The formal constraint in this case is heuristic rather than explanatory; it guides further inquiry by helping to narrow the space of plausible mechanistic models.

2.2. Searching for Mechanisms of Apoptotic Switching

Apoptosis is programmed cell death for multicellular organisms: nuclei and cytoplasm shrink, become encased in ‘apoptotic bodies,’ and send signals for being engulfed by nearby phagocytic cells. But apoptosis must be regulated. Too much promotes ischemic conditions and neurodegeneration; too little, cancer and autoimmune diseases. The principal intracellular components that regulate the mechanisms for apoptosis belong to the B-cell lymphoma 2 (Bcl-2) protein family. Some proteins in this family – the “anti-apoptotic” ones –inhibit apoptosis. Others – the BH3-only proteins – promote apoptosis. Among these promoter proteins, Bax and Bak are

especially important: they make permeable the mitochondrial outer membrane, and this mitochondrial outer membrane permeabilization (MOMP) makes cell death inevitable (Saikumar et al., 1999, 489-491).

There remains disagreement concerning how anti-apoptotic Bcl-2 proteins inhibit MOMP and thereby preserve cell viability, how BH3-only proteins facilitate MOMP and thereby promote cell death, and how the entire Bcl-2 switching network regulates Bax/Bak activation (Chen, Cui, Zhang, & Shen, 2007, 5143). This latter disagreement centers upon whether BH3-only proteins activate Bax/Bak directly, and how the anti-apoptotic Bcl-2 proteins inhibit Bax/Bak activation. Two kinds of solution are prominent. Some researchers favor a “direct” activation model, according to which BH3-only proteins directly activate Bax/Bak while anti-apoptotic proteins inhibit this activation by sequestering BH3-only proteins (see Galonek and Hardwick 2006). Others favor an “indirect” activation model, according to which BH3-only proteins activate Bax/Bak indirectly by virtue of binding and inhibiting the anti-apoptotic proteins, which themselves inhibit Bax/Bak activation by sequestering the Bax/Bak proteins directly (Youle, 2007).

Both models have some experimental evidence in their favor. Chen and colleagues tactfully summarize the evidential situation as one in which “proving which model is more telling by means of experimental approaches seems difficult thus far” (2007, 5143). But rather than recommend an agnostic attitude toward the models until better evidence is available, and rather than develop some kind of hybrid model, Chen and colleagues use what we call constraint-based reasoning in order to infer that the direct activation model is more plausible than the indirect activation model – and so, incidentally, to recommend that researchers concentrate their efforts on testing and improving the direct activation model. (For a reconstruction and analysis of their reasoning that is more detailed than what follows, see Jones, 2016)

Chen and colleagues introduce several measures of parametric robustness for judging the comparative plausibility of apoptotic switching models. These measures focus on a handful of features important to the apoptotic switch, all of which are exhibited by both direct and indirect activation models. These features, as named by Chen and colleagues, are:

Ultrasensitivity: Apoptotic stimuli produce small responses initially, and increasing the stimuli produces increasingly stronger responses (in comparison to a baseline model).

Range: There is some threshold such that, in non-apoptotic conditions, the basal activation level of Bax is below this threshold while, when apoptotic stimuli are present, the basal activation level for Bax is above this threshold.

Inhibition: There is some threshold such that, when apoptotic stimuli are present, the activity of Bax exceeds this threshold while, when anti-apoptotic proteins are present, the activity of Bax is below this threshold.

Insensitivity: There is relatively low Bax activity in non-apoptotic conditions, even with small perturbations of the apoptotic stimuli (Chen et al. 2007, 5145-5147).

Chen and colleagues use these features to define various measures of parametric robustness. Each involves a formal mathematical criteria for whether a model exhibits certain empirical features as well as for the number of different (suitably defined) parameter values for which the model exhibits those features. Chen and colleagues argue that most of their measures rank the direct activation model as more (parametrically) robust than the indirect activation model: “the direct activation model confers a better robustness in contrast to the indirect activation model” (Chen et al. 2007, 5148).. For example, they argue that the direct activation model is more parametrically robust than the indirect activation model with respect to the aforementioned group of four features because a simplified variant of the former model simultaneously exhibits all four features for 316 of 3000 randomly selected parameter sets, while a similarly simplified variant of the latter model does so for only 31 of 3000 sets.

Chen and colleagues use their arguments about the relative parametric robustness of competing models to rank the comparative plausibility of those models. In doing so, they endorse parametric robustness as a formal constraint upon plausibility. Doing so thereby allows them to use their ranking to narrow the search space of plausible apoptotic switching models. For the requirement that models exhibit a sufficiently high level of parametric robustness restricts the range of plausible models of the apoptotic switch from a space containing both direct and indirect activation models to a space containing only direct activation models. Tokar and Ulicny (2012) introduce certain “hybrid” models into this remaining possibility space, and they argue that similar robustness considerations eliminate these hybrid models as well.

2.3 The Principle of Constraint-Based Search

The case study in 2.2. illustrates how constraint-based reasoning helps researchers rank possible models with respect to their plausibility and thereby eases the task for finding relevant explanatory models. Formal constraints can narrow down the space of the possible by delineating (fallible and provisional) boundaries between plausible potential mechanisms and implausible ones. The notion of constraint-based reasoning draws on the idea behind the more specific strategy of constraint-based modeling. Constraint-based modeling has the following epistemic virtue that applies more generally to what we call constraint-based reasoning: “Given that biological information is typically incomplete, the idea of constraint-based modeling is to describe a biological system by a set of constraints, which characterize its possible behaviors, but in general do not allow to make a precise prediction” (Orth, Thiele, & Palsson, 2010). Figure 2 illustrates the idea of narrowing a solution space graphically. Without constraints to guide biological research, the space of possible models can be quite overwhelming.

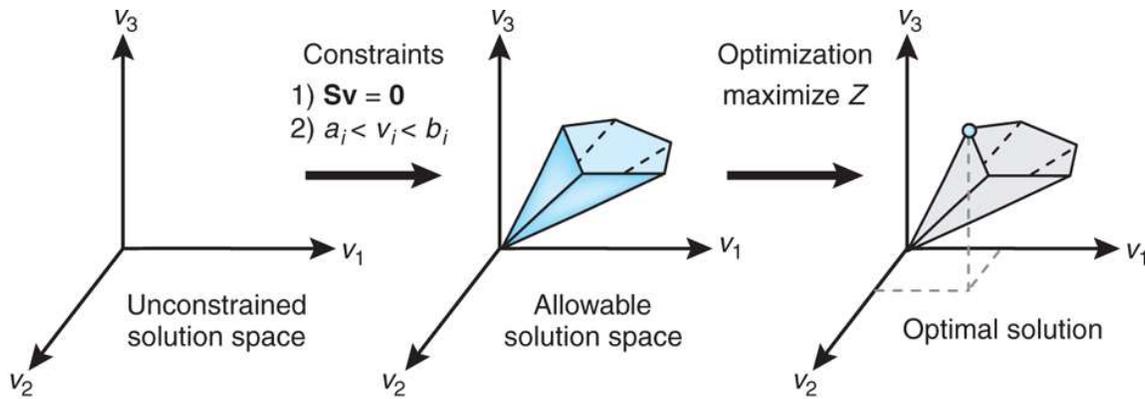


Figure 2. The principle of constraint-based search in an abstract solution space. The unconstrained space for this particular figure is supposed to range over possible flux distributions for a biological network; the first constraint is imposed by the stoichiometric matrix for this network, the second by upper and lower capacity bounds a_i and b_i ; optimizing the function Z determines a single optimal flux distribution on the boundary of the allowable possibility space. We intend the figure to illustrate the idea of using constraints to guide research and refer the reader to (Orth et al. 2010) for information about specifics of the visual. Reprinted by permission from Macmillan Publishers Ltd: *Nature Biotechnology*, Orth et al. What is flux balance analysis? Copyright (2010).

To further illustrate how formal constraints can work as research heuristics, we provide an additional brief example that also considers plausible models for biological robustness. Eldar and colleagues use formal constraints to delineate the space of possible models that can account for robustness of the bone morphogen gradient (BMP) in embryonic development of the fruit fly in *Drosophila* (Eldar et al., 2002). Initial experiments had shown that the BMP gradient is highly robust to genetic changes affecting gene dose of the constituents of developmental system, including the morphogen, inhibitors and other associated proteins. To narrow down the set of possible models, Eldar and colleagues formalize and simulate a large number of models, representing possible architectures of the network of interacting extracellular proteins that form complexes, diffuse to different locations of the developing embryo, and become degraded. They then use a formal definition of robustness as a constraint by setting a quantitative threshold value for possible models. Specifically, models with more than 10% change in BMP level, when reducing the gene dosages of the components by a factor two, were excluded. From over 66,000 simulations, varying nine network model parameters over four orders of magnitude, they find that only 198 network possibilities qualify for their formal definition of robustness. Strikingly, when examining the robust networks the researchers find that they all display a sharp concentration gradient of the morphogen, related to diffusion properties of the protein complexes. They use this discovery to generate the hypothesis that robustness is achieved through a “shuttling-degradation mechanism”, where an inhibitor acts as a shutting agent that facilitates the transport and storage of the active morphogen. These predictions have subsequently been investigated experimentally and confirmed.

What we take these examples from Chen’s and Eldar’s research groups to show is that defining and formalizing the constraints within which a given system must operate helps to narrow the possible solution space for identifying the specific causal processes currently maintained by a system. Whereas Chen and colleagues use a mathematical condition for

parametric robustness to judge the comparative plausibility of competing possible mechanisms of apoptotic switching, Eldar and colleagues use a mathematical condition for threshold shifts to generate a hypothesis about a likely mechanism related to embryonic patterning. In the coordinate system on the right in Figure 2, criteria for optimal solutions further constrain the search space. In the context of evolutionary biology, the heuristic of defining optimal solutions has typically been considered an adaptationist strategy. However, optimality can serve, instead, as merely a search tool for a possible testable solution to a mechanism operating in extant organisms (for further discussion on this issue, see Green, Levy, & Bechtel, 2015).

We believe that our account to be compatible with much of the literature on heuristics in science. Several recent accounts address the role of mathematical abstractions as model templates, e.g. Humphreys' (2004) account of computational templates, and through the notions of mechanism sketches or schemas (Craver, 2007; Darden, 2002). For these accounts, abstract formalizations can be transferred as templates from other similar systems and thereby guide the search for core causal factors. This way, templates can serve as starting points for more developed models of concrete systems. Templates or schemas also can be applied in an exploratory fashion across multiple domains, because they are more abstract and therefore have a greater scope of applicability compared to more detailed mechanisms. Templates help to rule out possibilities incompatible with, or made improbable by, conditions imposed by the template; and templates facilitate the goal of identifying more-or-less detailed explanatory mechanisms about sequences of causal processes. But whereas the heuristic role of mathematical abstractions is generally accepted, the *explanatory* status of formal constraints in biology remains controversial. Some mechanistic accounts view mathematical abstractions merely as offering predictive and descriptive "how-possibly" models rather than genuine explanatory roles (e.g., Craver, 2007; Matthiessen 2015). In the following section, we argue that restricting abstractions to a role as tools for identifying causal specifics overlooks important explanatory roles of formal constraints. We investigate the explanatory significance of formal constraints, in particular their variable importance for dynamic mechanistic explanations and constraint-based explanations.

3. Design principles of biological robustness: explaining via formal constraints

Proponents of so-called *dynamical mechanistic accounts* have argued that dynamic models and other mathematical abstractions can play an explanatory role in the life sciences (Bechtel & Abrahamsen, 2012; Brigandt, 2013; Levy & Bechtel, 2013). Dynamical mechanistic explanations emphasize organizational and dynamical aspects of mechanisms. Levy and Bechtel (2013), moreover, emphasize that mechanistic explanations can operate on different levels of abstraction and that the search for general patterns of causal connectivity is perfectly compatible with mechanistic accounts. To illustrate the idea, we provide a brief example.

Tyson (1991) uses biochemical evidence to postulate a potential mechanism for cell cycle oscillations in early *Xenopus* embryos, consisting of a handful of mechanism components and interrelations. He constructs from this mechanism a mathematical model of cell cycle oscillations, as "a reasonable 'first approximation' to the cell-cycle regulatory network" (1991, 7329). This construction appeals to various formal constraints. For example, Tyson applies a (phenomenological) mass-action rate law – in the form of a mathematical equation – to the

reaction relations in his posited mechanism in order to obtain kinetic equations for the changes in various biochemical species over time (see Tyson 1991, 7329; Jones and Wolkenhauer 2012, 714-716). Tyson then uses the model he obtains from his posited mechanism to show, via mathematical analysis, that the mechanism for the cell cycle can operate in three modes. Because each of these modes corresponds to an experimentally well-attested mode of the cell cycle, Tyson's analysis demonstrates that his posited mechanism is organized in a way that is capable of producing experimentally observed phenomena. His modeling thereby provides a dynamic mechanistic explanation for certain phenomena related to cell cycle oscillations in early *Xenopus* embryos.

For dynamic mechanistic explanations, formal constraints help to clarify how the *organizational aspects* of specific mechanisms influence the mechanisms' behavioral manifestations. Importantly, the explanatory power of these explanations rests on the integration of abstract models with mechanistic details. We intend the notion of constraint-based explanation to highlight, instead, the prevalence of explanatory contexts that aim to clarify *why* a structural or dynamic pattern occurs in systems with heterogeneous mechanistic detail. This often involves determining the extent to which system properties depend on specific causal details about the properties of the components that make up the system or on general organizational features. This raises the question of whether reasoning about the *generality* of the organizational structures in different causal contexts should be characterized as mechanistic, or whether researchers draw on different explanatory strategies for different purpose. We adopt the latter stance, and we offer in support considerations about design principles and so-called design explanations.

Green and Wolkenhauer (2013) have argued that organizing or design principles play two important roles in biological research. One is akin to the heuristic role outlined in Section 2: to facilitate transfer of resources across research contexts by serving as general and abstract templates for discovering similar mechanisms in other systems. The second role has received less attention from philosophers of biology. Design principles play an important theoretical role in identifying rules, principles and laws for which some functioning or organizational scheme is an instantiation. To reason about design principles is not only to reason about the workings of specific real-world systems but to identify the categories of abstraction that these belong to. Accordingly, reasoning about design principles often involves a reflection about whether certain system architectures are necessary or sufficient for a certain function to be performed, and such reasoning facilitates understanding why certain dynamic or structural properties are present. For instance, the functional dependencies associated with abstract schemes, such as network motif diagrams, identify general features of a larger class of regulatory circuits and thereby categorize how specific mechanisms relate to general, and more abstract, types such as amplifiers, persistence detectors etc. (Alon, 2007). Although categorizing these types is fruitful for, and related to, mechanistic research, the notion of design principle implies a role for unification not typically associated with mechanistic accounts (see also Section 5). Design principles categorize seemingly different biological processes into types through a demonstration of a general principle that they all instantiate, and this categorization makes many causal details or differences between biological processes irrelevant for understanding certain system behaviors. Importantly, unification in this context does not mean reduction of mechanistic explanations to principles. Rather unification is reached through higher-level reflections on *types of system organization*.

Accordingly, the virtue of design principles lies not only in their heuristic value as templates for discovering similar mechanisms in other system, but *also* in their role for theorizing about what certain types of systems have in common and why.

We can clarify this explanatory role of design principles, and illustrate how formal constraints sometimes carry more explanatory weight than mechanistic details, with an example. The example concerns an analysis by John Doyle's system biology group to identify the design principle behind the perfect adaptation of bacterial chemotaxis (Yi, Huang, Simon, & Doyle, 2000).⁴ Bacterial chemotaxis is the strategy by which bacteria can "swim" toward attractants and away from repellants via flagella motion. Flagellated bacteria move in a medium via a series of straight-line runs followed by "tumbles" that randomly change their directions. Addition of a chemoattractant to the bacteria's environment changes the frequency of such tumbles and the bacteria conduct a biased random walk. Experimental studies have showed that the system of sensors and flagella motion display robust perfect adaptation (RPA) to the level of nutrients such that the system remain sensitive to changes regardless of changes in the average level of attractants or repellants. In other words, if the frequency of the flagellum ceases due to a nutrient, the system will return to its prestimulus value (steady-state tumbling frequency) by reactivating the receptor, even if the signal is still present. Adaptation thus maintains sensitivity of sensors over a wide range of system parameters, rendering the steady-state behavior independent of specific concentrations of attractants or repellants in the environment.

The molecular mechanisms behind the system are well understood. Chemotaxis involves mediation of a kinase-phosphorylation signal transduction pathway from methyl-accepting transmembrane receptors to six Che proteins that interact with the flagella motor. Adaptation to concentration of attractant and repellants is causally instantiated by methylation and demethylation of receptors (Barkai & Leibler, 1997; Yi et al., 2000). Mechanistic explanations explain how interactions among the aforementioned components causally produce adaptative responses. Hauri and Ross (1995), for example, construct a set of differential equations using mechanistic details about *E. coli's* chemotaxis network; then, using simulations, they define a range of parameter values for which these equations exhibit RPA. If the aim is a realistic model of a specific system, further specification and de-idealization of a model would typically increases the explanatory power of the model. But we wish to draw attention to contexts in which 'how-actually' explanations for specific causal processes are not considered the end point.

A different question one may ask is 'how-possibly' the same function can be realized in different systems, and about the extent to which the relations between structures and functions are dependent on specific parameter values or a general design principle. In the context of perfect adaptation in biological systems, one question of interest for engineering-inspired systems biologists is whether the *E. coli* chemotactic network and other similar systems exhibits RPA because of some general design principle of its network structure. Doyle's research group approached this question by reinterpreting the Barkai-Leibler model in the mathematical framework of control theory. In our view, what they do is to provide a constraint-based

⁴ The example is mentioned in (Braillard, 2010) as an example of a design explanation but we here develop the example in further detail in the broader context of constraint-based explanations. Section 4 discusses how these are related.

explanation for why many bacterial systems, despite their differing mechanistic details, exhibit perfect adaptation (Yi et al. 2000).

A common engineering goal is to ensure that the performance of a system is robust to noisy input signals or system parameters. To design systems that asymptotically track a fixed steady-state value, engineers rely on the design principle of integral feedback control. The strategy of Doyle's group was to explore whether the same principle i) could account for robust perfect adaptation in chemotaxis in the previously developed Barkai-Leibler model, and ii) whether this principle is necessary for the phenomenon. In integral feedback control, the difference between the actual output (y_1 : the actual receptor activity) and the desired output (y_0 : steady-state receptor activity) is fed back to the system as the integral of the system error on x ($-\int$ on Figure 3). The principle ensures that the process with gain k and input u (chemoattractant) is normalized to provide output y by accounting for the difference between y_1 and y_0 .

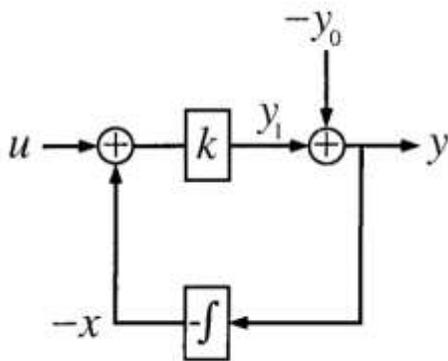


Figure 3. A diagram illustrating integral feedback control. The k -box represents a system process, which receives input u from outside the system as well as additional feedback input, for a total gain k . The rightmost plus sign in the circle represents the system comparing the actual output y_1 from the k -box process with a desired steady-state output y_0 . The integral-box represents the feedback controller, which receives the error y from the prior comparison and feeds a time integral of this error back into the system as feedback input.

Source: Yi TM, Huang Y, Simon MI, Doyle J (2000) Robust perfect adaptation in bacterial chemotaxis through integral feedback control. Proc Natl Acad Sci USA 97:4649-4653. Copyright (2000) National Academy of Sciences, USA.

Drawing on control and dynamical systems theory, Doyle's group abstracts from what the specific parameters in the Barkai-Leibler model refer to and derive a general equation from the model that is known to be characteristic for integral feedback control in engineering. The rigorous mathematical framework of control theory allows the group to systematically test the dependence of RPA on specific the steady-state solutions and assumptions of the Barkai-Leibler model. The equation guarantees that if the concentration of bound Che R protein to the receptor complex is independent of ligand, adaptation will be independent of the concentration levels of the attractants. Furthermore, the group demonstrates mathematically that only four assumptions of the Barkai-Leibler model (one of several mechanistically detailed models for bacterial chemotaxis) are needed for integral control. The assumptions are as follows: CheB demethylates only active receptors; kinetic rate constants of CheR and CheB are relatively independent of methylation state and ligand occupancy of receptor complex; activity of unmethylated receptor is negligible relative to methylated receptor forms; concentration of bound CheR is independent of ligand level.

The next step is to explore whether integral feedback control is *necessary* for robust perfect adaptation. If so, the equation derived would be a formal constraint on any possible model of a system with this property.⁵ Doyle and colleagues prove that any linear system displaying robust asymptotic tracking must be linearized around a fixed ligand equilibrium described by the equation derived for integral feedback control (see appendix of Yi et al. 2000). They generalize this finding also for non-linear systems in arguing that this is a special case of the internal model principle (IMP) in control theory. Thus, Doyle's group considers integral feedback control to be both necessary and sufficient for robust perfect adaptation (perfect adaptation not sensitive to specific parameter values). In their words,

When a system exhibits robust asymptotic tracking [perfect adaptation], it must have integral feedback as a structural property of the system. When combined with biological realizability, this may greatly constrain, on the basis of external behavior, the possible internal mechanisms that can be used to achieve the observed behavior (Yi et al. 2000, 4651).

Thus, whereas mechanistic explanations have the affordance of describing how adaptation in bacterial chemotaxis is causally produced in concrete biological systems, Doyle's group aims at answering a question about what all these systems must have in common (for a review of more recent work of this type, see Iglesias, 2013). Mechanistic explanations can capture the fine-grained differences between biological systems by describing, separately, how a dynamic pattern is causally instantiated through interactions between molecular components. Constraint-based explanations are too abstract to capture such differences. Yet, abstraction has the complementary value of specifying more generally what is *possible* or *probable*. Constraint-based explanation complements these explanations by explaining *why* the two systems exhibit the same pattern, despite their causal details.

An implication of the abovementioned analysis is that for answering some explanatory questions, many causal details become irrelevant. Placing the Barkai-Leibler model in the general framework of control engineering reveals that the design principle – integral feedback control – is independent of the other assumptions of the Barkai-Leibler and serves as a formal constraint on any model of a system displaying robust perfect adaptation:

[I]ntegral control is not only sufficient but also necessary for robust perfect adaptation. Thus, if their specific model is later found to be contradicted by experimental data, another mechanism implementing integral feedback is likely to be present (Yi et al. 2000, 4652).

In other words, even if Barkai and Leibler's mechanistic model is revised over time, or if a model is developed for different bacterial systems, one can expect the design principle to remain the same. Furthermore, the researchers expect the design principle to be found across systems with very different causal details, from biological systems to jet airplanes and transistors (see also Braillard, 2010). The principle not only helps as a search strategy for narrowing down the scope of possible mechanisms for other systems, but also to explain why the same principle is employed in so

⁵ Perfect adaptation is also found in other biological systems, e.g. calcium homeostasis in mammals, membrane turgor pressure in yeast etc. This way, the findings of Doyle's group has a much broader target than bacterial chemotaxis in *E. coli*.

different systems. Thus, their approach de-emphasizes the importance of mechanistic detail and prioritizes the importance of formal constraints.

So far, we have not evaluated whether Doyle and colleagues succeed in proving that integral control via negative feedback is necessary. Recently, other researchers have argued that a network structure with an incoherent feedforward loop architecture can realize the same function (Ma et al. 2009). But, interestingly, they argue that these two design principles are the only possible topologies. Moreover, a recent review argues that both are instances of integral control that exemplifies how the internal model principle is satisfied in all known biological cases (Iglesias, 2013). What we want to emphasize is that these researchers engage in explanatory projects that do not aim to provide causal explanations of particular systems. Ma and colleagues articulate the aim as follows (2009, 761): “Here, instead of focusing on one specific signaling system that shows adaptation, we ask a more general question: what are all network topologies that are capable of robust adaptation?” The example highlights a type of analysis that serves to “unify the organization of diverse circuits across all organisms” (Ma et al. 2009, 760). Thus, this type of analysis addresses questions about the scope of biological variation and about constraints that result in characteristic biological patterns.

In summary, while the possibility space for constraints used as heuristics ranges over potential mechanisms, the possibility space for constraint-based explanations ranges over potential variations among structural aspects of biological systems, given specified functional requirements. Philosophers of biology have given little attention to the importance of such abstraction-based explanatory strategies in biology. But a number of recent philosophical papers are exceptions in pointing to explanatory strategies in biological research that differ from mechanistic ones as typically conceived. In the following, we introduce the term of constraint-based explanation to highlight common features of recently developed non-mechanistic accounts.

4. Constraint-Based Explanation in Biology

The ideal of what we call constraint-based explanation is longstanding, dating back at least to D’Arcy Thompson’s (1917/2004) work on mathematical dependencies between growth and form and Ludwig von Bertalanffy’s (1969) *General System Theory*.⁶ Yet, modern research on biological robustness is making increasingly apparent the theoretical motivation for focusing on abstract generalizations. For example, in systems biology, a key question type is whether a regulatory circuit in a metabolic network functions as a persistence detector, pulse generator, switch, amplifier, or other abstract mechanism by virtue of its architectural design rather than its specific molecular components (Alon, 2007). Biologists uncover such features not only by varying conditions within experimental settings in order to investigate which parameters make a difference, but also by analyzing the architectural features in mathematical formalizations of the systems’ causal interaction networks. This latter research aims to understand not only how a function is causally produced, but also the degree to which a property depends upon specific

⁶ For a brief historical review of the quest for general principles in biology, see (Green and Wolkenhauer, 2013).

constituents or parameter values, and the ways in which this property is insensitive to their specific causal constitution.

For example, as we noted in Section 2, a number of recent research papers indicate that topological features of biological systems often ensure robustness of functions across a range of perturbations to lower-level constituents and environmental conditions. The explanatory aim for constraint-based explanations is to outline general relations between organizational features and system behavior. These explanations aim to understand not only how macroscale behaviors emerge from microscale structures but also the extent to which the macroscale behaviors are insensitive to variations of specific microscale details. In biological research, the use of formal constraints towards this aim is often conceptualized as a search for design or organizing principles – i.e., formalizations connecting well-defined properties of biological networks to functional properties of a mathematically defined system (Hornberg et al., 2005; Shinar & Feinberg, 2011; Wolkenhauer, Shibata, & Mesarović, 2012). Other explanations aim to understand qualitative dispositional behaviors – such as vulnerability to specific kinds of attacks but widespread resistance to others – in terms of large-scale organizational features common to many different networks. Formal constraints relevant to this aim ascribe certain topological properties to biological networks (Huneman, 2010). We return to these examples in Section 4.2. First, however, we consider in more detail the notion of constraint-based explanation and its relation to dynamic mechanistic explanation.

4.1. Template for Constraint-Based Explanation

We begin by clarifying the difference between dynamic mechanistic explanations and constraint-based explanations. Dynamic mechanistic explanations use formal constraints to model mechanism operations, and in turn use these models to explain phenomena as effects of certain mechanism behaviors.

Dynamic mechanistic explanations typically involve the following ingredients:

1. A claim that some system S embodies some mechanism M with certain specific components and some specific organizational relation among those components.
2. A demonstration that formal constraints F , together with M , define a computational model that exhibits some behavior B .
3. An inference that S exhibits B by virtue of the organization and interactions of the components of M under constraints F .

Dynamic mechanistic explanations are driven by the aim to understand the causal functioning of concrete biological systems, in part from the way these are organized and in part as a result of properties of the components. To be explanatory, a model in this framework must integrate organizational or dynamic schemes with mechanistic details. So, for example, Tyson's explanation of certain cell cycle behaviors (Section 3) depends upon his posited mechanism being relatively accurate with respect to how certain biochemical species relate to each other. He combines his posited mechanism with formal constraints such as the law of mass action in order to define a

computational model of the cell cycle. However, the formal constraints used to construct mathematical models from posited mechanisms need not be accurate or realistic so long as the behaviors of the models they help to approximate the actual behavioral manifestations of the posted mechanisms.

Constraint-based explanations, unlike dynamic mechanistic ones, do not give as much explanatory weight to mechanistic details about properties of system components. They involve, instead, the following ingredients:

1. A claim that systems within some class **C**, differing with respect to some range of mechanistic detail, share some organizational or structural property **O** and some dynamic or functional property **D**.
2. A demonstration that formal constraints **F**, applicable to systems within **C** despite their heterogeneous mechanistic detail, define or limit the range of possible dependence relations between **O** and **D**.
3. An inference that all systems in **C** exhibit the permitted **O-D** dependency relations, regardless of differences of mechanistic detail among members of **C**.

The first ingredient of a constraint-based explanation involves generic abstraction in specifying structural or functional features that a set of systems (with differing mechanistic details) have in common. The second ingredient explains this correlation, and the irrelevance of certain mechanistic details to the correlation, with reference to a shared set of (formally defined) constraints. The third step is an investigation of the extent to which all systems of this class must obey such constraints and, accordingly, share the permitted **O-D** dependency relation. Thus, this explanation is driven primarily by the wish to understand the generic features of a class of systems shaped by a shared set of constraints. The formal constraints not only show why that specific system exhibits the phenomenon of interest, but also why other systems, with some different mechanistic details but with the same system organization, would exhibit the same phenomenon.

As an illustration of the template, recall the example where Doyle's group demonstrates that integral feedback control not only is sufficient but also necessary for robust perfect adaptation. They explicitly state that even if the specific mechanistic model turns out to be incorrect, the revision would regard the mechanisms implementing integral feedback but not the principle as such (Yi et al. 2000, 4652). This type of explanation thus provides a unifying framework at a more abstract level for the type of mechanisms that one may expect to find. Yet, unification in this context does not entail reduction of specific explanations to a general scheme. Rather, the different types of explanation answer different questions – about how a phenomenon is causally produced *or* about the possible design principles systems of this type can or must instantiate. Accordingly, the weighting of mechanistic details and formal constraints, respectively, varies for the two types of explanation (as illustrated on Figure 4).

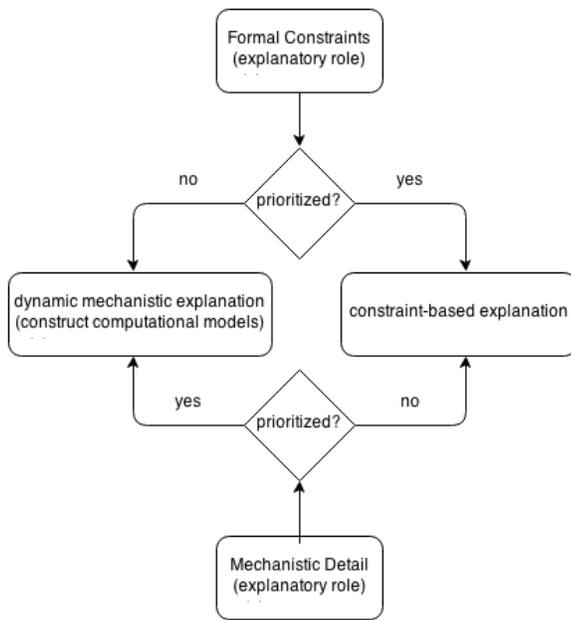


Figure 4: Explanatory salience expressed as different explanatory priorities for formal constraints and mechanistic details.

To argue that formal constraints are explanatory salient is uncommon in the context of philosophy of biology. Yet, examples of this type of research are becoming more and more prominent, especially in biological research on robustness.⁷ It is sometimes said that there are no laws or theorems in biology. Nevertheless, theorems have recently been used to uncover biologically relevant features in systems biology. An example is Shinar and Feinberg's (2010, 2011) robustness theorem for biochemical reaction networks. Drawing on Chemical Reaction Network Theory, their mathematical proof illuminates how certain architectural constraints of biochemical networks guarantee the robustness of one biochemical species against changes in the other network constituents. More specifically, they demonstrate that any mass-action network, regardless of its size, displays absolute concentration robustness (ACR) provided that the deficiency of the network is one and there are two non-terminal nodes that differ only in one biochemical species. These provisions of the theorem are what we have been calling formal constraints. Deficiency, for example, is a measure for the interdependency of the reactions in the network, calculated by subtracting the linkage class and rank of the network (the number of independent reactions) from the total number of nodes. The robustness theorem demonstrates that any mass-action network systems to which these constraints are applicable must display a certain kind of structural-functional dependency, namely, ACR. They further show why some network designs cannot display such robustness.

Shinar and Feinberg's approach to network robustness is purely theoretical. Their proof makes only very minimal assumptions about the mechanistic structure of systems exhibiting ACR. Nonetheless, their result points toward relevant properties that might unite highly robust

⁷ See, for instance, Woodward's discussion of whether abstract research strategies on the robustness of the mitotic spindle assembly system can be captured in a mechanistic framework (Woodward, 2013).

biological networks. So far, the conditions of the theorem have been identified experimentally in two systems in *E. coli* which exhibit approximate ACR. Shinar and Feinberg's theorem, and other associated results, suggest that the robustness of these systems might not reflect fine-tuning of their parameters or the existence of specific mechanisms but, instead, general architectural features shared by mechanisms with heterogeneous detail.

4.2. Unifying Non-Mechanistic Explanations as Constraint-Based Explanations

Recent philosophical literature highlights several alternatives to mechanistic explanatory strategies. We have mentioned already explanation via design principles. But there are also non-causal physical explanations (Morange, 2005; 2011), dynamical explanations (Berger, 1998), type II or minimalist model explanations (Batterman, 2002; Batterman & Rice, 2014; Gross, 2015), optimality explanations (Rice, 2012), mathematical explanations inspired from dynamical systems theory in neuroscience (Ross, 2015), and principles of self-organization (Kauffman, 1969). It is interesting to note that, despite variations among different types of explanations, several authors compare the explanatory virtue of their abstract accounts to Sober's (1983) equilibrium explanations to illustrate how explanatory power in some cases can be dissociated from causal details (c.f. Batterman & Rice, 2014; Braillard, 2010; Gross, 2015; Hamilton, 2007).⁸ Similarly, these strategies share with constraint-based explanation abstraction from causal and mechanistic detail and draw on tools from nonbiological fields, such as engineering, mathematics, or physics.

Our modest goal in the following sections is to situate two of these alternatives – design explanations and topological explanations – within our template for constraint-based explanation. Importantly, we do not wish to disregard the value of the fine-grained differences among these explanatory strategies.⁹ Just like the label 'mechanistic explanations' is useful for capturing commonalities between variants of mechanistic accounts held by philosophers, so is it our hope that the notion of constraint-based explanations can help clarify central features of the accounts of explanation mentioned above. These explanations are *prima-facie* quite diverse. Design explanations typically focus upon macroscale structure-function dependencies, while topological explanations typically focus upon dependencies between microscale structure and macroscale behavior. Unlike topological explanations, design explanations also tend to be contrastive, comparing actual organism "designs" with a range of alternatives. Situating both kinds of explanation within our template for constraint-based explanation shows that, despite these superficial differences, there is an underlying commonality to the explanatory strategies beyond their disregard for mechanistic detail, namely, their emphasis on formal constraints.

⁸ Sober's key example of equilibrium explanations is Fisher's mathematical model of sex ratio distributions. The model explains why the sex ratio would be the same regardless of which causal scenarios preceded the observed event.

⁹ We also do not wish to disregard potential connections between constraint-based explanation and the growing literature on mathematical explanations in the natural sciences, as represented in works such as (Batterman, 2010; Pincock 2011, and Lange, 2015). This literature tends to focus upon the distinctive roles of mathematics in scientific explanations and how mathematical content manages to apply to the empirical world. Our project here, by contrast, is narrower in several respects. We focus upon mathematical content only insofar as it functions as a formal constraint in biological science, rather than mathematics more broadly construed and across the natural sciences. We also focus upon how these constraints complement mechanistic search and explanation, rather than how the constraints apply to the world in a more general sense. For discussions similar in focus to ours, see (Braillard and Malaterre, 2015).

4.2.1 Design Explanation

Wouters (2007) coined the term ‘design explanation’ to highlight the important aim in proximate biology of determining “the constraints on what can be alive”. Design explanations clarify why some forms or traits are not possible given that organisms with certain characteristics must live in environments posing certain challenges for survival, and thereby why certain design patterns are observed across lineages of common descent. Design accounts do not provide historical explanations of conservative or convergent evolution, but instead examine the scope of *possible* or optimal structures that could realize a certain function. For instance, the possibilities for functional respiratory designs are highly limited once biologists take into account constraints such as rates of gas diffusion, energy requirements for animals of different sizes and accessibility to oxygen in water and land environments. Similarly, biologists as well as engineers may be interested in delineating the scope of wing shapes that would allow a plane or an organism to fly, or which organizational network schemes lead to sustained oscillations (Jaeger and Sharpe, 2014). Design explanations aim to understand the *scope* of possible or good designs, rather than to reverse engineer the mechanisms behind some specific observed design.

Design explanations are contrastive, comparing design options with respect to their usefulness for staying alive given certain structural or functional constraints. These comparisons often involve demonstrating that certain structure-function pairings are impossible, and they do so by pointing to invariant generalizations such as Fick’s law of gas diffusion, gravity, geometrical scaling, biochemical constraints for energy consumption etc. But design explanations are not reducible to these generalizations. Generally, design explanations (or sentential representations thereof) involve the following ingredients:

1. A claim that organisms within some class **C**, differing with respect to certain biological details, share certain trait or “design” **T** with functional property **P**.
2. A demonstration that certain invariant generalizations, principles, and other formal constraints **F** limit the scope of possible designs for **C**-type organisms’ exhibiting **P**.
3. An inference that there are formal dependency relations between **T** and **F**, such that having **T** is better for an organism’s in **C** exhibiting **P** than is having alternative traits Z_1, \dots, Z_n , regardless of differences of biological detail among members of **C**.

These ingredients are straightforwardly instances of the ingredients given by our template for constraint-based explanation. Taken together, they explain why organisms within the relevant class have the better-designed trait **T** for performing function **P**. The comparison traits Z_1, \dots, Z_n are conceivable designs, and they need not be instantiated in real systems in order to be options. Given these design options, the class-wide function **P**, together with various other claims **F**, supports counterfactual comparisons about design superiority. But design explanations do not reduce to these other claims, because they involve understanding the functioning of a system as a whole in a particular context. Whence, Wouters describes design principles as non-causal counterparts to so-called ‘biological role functions’. The latter explains how the orchestrated organization of a system’s components causally contributes to the capacities of a system

instantiating such functions. Like biological role functions, design explanations are not evolutionary or historical. But unlike biological role functions, the functional dependence between T and P does not refer to an asymmetric process in time between cause and effect. Rather, the relations are symmetric dependencies between variables, similar to the relation between volume and temperature in the ideal gas law. Moreover, the explanation does not show how, or depend upon whether, the candidate design traits are causally or historically instantiated in specific organisms.

In summary, design explanations do not point to causal processes or historical developments. Instead, they explain why a certain design is possible or preferential compared to alternatives. Braillard (2010) clarifies the close association between this type of explanation and the quest for so-called design principles in systems biology. Design principles describe structure-function relations in organisms or artifacts that are common due to a shared set of constraints. Braillard points to how they also categorize what is possible at all, making an analogy to electronic design: “[T]here are many ways to build an amplifier, some good and some bad, but general principles of signal amplification must nevertheless be respected if the amplifier is to work at all” (Braillard, 2010). Thus, from a general perspective, the function of a system determines the boundaries of the space of possible structures and vice versa. Principles of design, accordingly, point toward general schemes that, on a more abstract level, unite specific instantiations under a common abstract and organizational pattern.

4.2.2 Topological Explanation

Design explanations often develop in close connection to engineering approaches to biology. Approaches to biology that invoke network and graph theory exemplify another kind of non-mechanistic explanation. Huneman (2010) describes these accounts as ‘topological explanations’. Network approaches to biological systems prominently feature two kinds of mathematical structures. Topological structures deal with connectivity properties of the state space used to represent system processes, while dynamical structures deal with defined relations among the state variables used to represent the system (Oster & Desoer, 1971, 221). Topological approaches are fruitful when parameter estimation and other difficulties hinder dynamical approaches (Ross, 2015). They are particularly salient in recent graph-theoretical approaches to uncovering the implications of common network structures such as scale-free, small world, hierarchical topologies, as well as overabundant circuits called network motifs. Topological approaches to explanation are increasingly common in fields like neuroscience, sociology, systems biology, and ecology, where extensive use of network representations reflects an interest for understanding how structural features (positively and negatively) constrain possible behaviors of the system through distributed functionality and robustness.

Topological explanations identify generic topological properties of living systems. While abstracting away from the causal basis for those properties, they emphasize features of network architecture that are invariant across a wide array of causal changes (Jones, 2014). These topological properties do not depend upon specific mechanisms or causal interactions. They concern, instead, how a system’s features – such as its physical components or mathematical regions in a phase space – relate to each other, and which of various system perturbations do not

affect those relations. Generally, topological explanations (or sentential representations thereof) involve the following ingredients:

1. A claim that systems within some class **C**, differing in certain microscale details, have underlying network structures that exhibit some microscale topological property **T** and macroscale dynamic capacity **D**.
2. A demonstration that formal constraints **F** entail that there are certain dependencies between **T** and **D** for systems in **C**.
3. An inference to the claim that systems in **C** exhibit these T-D dependencies, regardless of the systems' microscale differences.

Also in this context, the ingredients are straightforwardly instances of the ingredients given by our template for constraint-based explanation. They produce topological explanations that allow researchers to draw inferences, either of dynamic capacities from an analysis of topological features or, in the opposite direction, from capacities to the necessary topological features able to support them. These explanations emphasize formal constraints, rather than causal details of specific systems, as key to questions about general dependency relations between topology and function.

Kitano and Oda's explanation for a particular fragility of the human immune system nicely illustrates the notion of topological explanation. They write,

The adaptive immune system exhibits a clear bow-tie structure at both the cellular interaction and signal transduction levels. ... Owing to the structure of the intercellular interactions that form the bow-tie structure, with naïve CD4+ T-cells as the core of the network, the system is vulnerable to attack on CD4+ helper T-cells (Kitano & Oda, 2006, 2).

The bowtie structure to which they refer is a topological property of the immune system, and many diverse systems – including the world wide web – instantiate the same structure (Jones, 2014). That bowtie structures are vulnerable to core-centered attacks is a mathematically specified consequence of the network architecture. More generally, the relations between topological property and system capacity are not causal or evolutionary. Even though there is a dependency between topology and dynamic disposition, the dependence is structural rather than temporal or historical, and establishing the relation does not involve attending to how either explanatory element is causally or historically realized in specific systems.

5 The Autonomy of Constraint-Based Explanations

We have argued that both design and topological explanations are species of an explanatory genus that we call constraint-based explanation. For design explanations, the formal constraints concern structural or functional properties of organisms, and these constraints determine dependency relations between traits and functions. For topological explanations, the formal constraints determine dependency relations between topological network structures and dynamic network

capacities. In both cases, the explanatorily salient are formal dependence relations rather than mechanistic details. We conjecture that many other kinds of non-mechanistic explanation invoke formal constraints in similar ways. But rather than attempt to further establish the extent to which the focus on formal constraints unifies otherwise diverse kinds of non-mechanistic explanation, we turn in this section to the issue of whether constraint-based explanations are autonomous from, or parasitic upon, mechanistic explanations.

Our view is that emphasizing the explanatory force of formal constraints does not diminish the importance of mechanistic explanations. In fact, the different research strategies – one focusing on identifying formal constraints and determining their significance, the other on identifying mechanistic details and determining their significance – are often mutually supportive. We often need to know a great amount of causal detail before we can determine the extent to which system-level properties depend upon changes at this level (Green, Fagan, & Jaeger, 2015). Similarly, mechanistic explanations often must rely on information regarding more general structural and dynamic constraints in order to be fully explanatory. Despite this symbiosis, where constraint-based explanations may play a role in mechanistic explanations, we maintain that constraint-based explanations are not just a species of mechanistic explanation.

In our view, the dynamical mechanistic account correctly clarifies how understanding mechanisms and understanding abstract organizational schemes are often mutually dependent and supportive. Huneman (2010) calls such cases in which mechanistic and topological explanations support each other to produce a full explanation for ‘constraining topological explanation’. Huneman maintains, however, that distinguishing between mechanistic and (constraint-based) topological explanations is fruitful for understanding cases where topological properties take center explanatory stage and mechanistic details on the properties of part provide only limited explanatory support. Similarly, we have argued that constraint-based explanations are distinct from dynamical mechanistic explanations. We have argued, in previous sections, that constraint-based explanations do not explain how phenomena are causally produced through sequences of events but, instead, explain by identifying (typically formal) dependencies between different factors. The value of constraint-based explanations in biology clearly depends upon systems that causally realize the constraints and patterns they aim to describe. Yet, constraint-explanations do not point toward temporal sequences of causal events, as mechanistic explanations typically do, and they are not confined to structures or patterns that are causally instantiated in concrete biological systems. Whereas mechanistic explanations are typically seen as improved through steps that de-idealize the model, adding more details would, for constraint-based explanation, often be considered as ‘explanatory noise’ that makes the explanation context-dependent (see e.g., Batterman, 2002; Ross, 2015). Unlike many mechanistic explanations, constraint-based explanations often look the same for different causal systems (e.g., living and engineered). More important, however, is that constraint-based explanations refer to formal dependency relations that are explanatory by virtue of identifying *boundaries* for causal explanations. In other words, constraint-based explanations delineate a larger, but constrained, space of possibilities whereas mechanistic explanations reveal how operations are causally instantiated in concrete systems.

Whether non-causal or non-mechanistic accounts identify genuine explanations remains controversial among many philosophers of biology. In our view, philosophical accounts of explanation should pay attention to different explanatory practices in science rather than seek an overarching one-size-fits-all account. We believe that attention to different explanatory aims and strategies can better make sense of scientific practice, particularly of challenges and opportunities arising in interdisciplinary settings where different explanatory frameworks are combined. For instance, debates within the life sciences often turn on disagreements regarding the relative importance of formal constraints as opposed to mechanistic detail. One example is how experimental biologists often ignore the work of dynamical systems (DS) theorists because it is too theoretical and lacking either empirical grounding or experimental implications (Fagan, 2015). The differences between these explanatory standards can impede collaboration between theoretical and experimental biologists, but also in some cases lead to new innovations (Green, Fagan & Jaeger, 2015). The problem we see for the monistic accounts of explanations is that they fail to account for the existence of such disagreements and innovative potentials when different explanatory standards meet.

Science is complex and explanations answer a range of different questions. We have argued that there are other explanatory aims in biology than specifying how a specific phenomenon is causally produced. With the notion of constraint-based explanations we point to a class of non-mechanistic explanations that answer different explanatory questions about the boundaries of the possible biological variation and about why general models can often be applied to causally different systems. These questions cannot be answered via extrapolation from specific models alone but require generic abstraction through which mechanistic details are suppressed for the sake of defining general dependency-relations. Due to these important differences, categorizing all biological explanations under the same headline would, in our view, underestimate the richness of explanatory aims in the life sciences.

6. Concluding Remarks

In this paper, we have aimed to clarify how, for both heuristic and explanatory purposes, constraint-based reasoning strategies help to define the boundaries for cause-oriented strategies. Given a description of the formal constraints that a specific causal process must obey, researchers can draw on constraints to explore the hypothesis space for a given biological mechanism and thereby narrow down the spectrum of possible models to investigate experimentally. In addition to this heuristic role, we argue that constraint-based reasoning also serves an explanatory role in biology. Constraint-based explanation, as defined in this paper, is an abstract explanatory strategy that seeks to uncover how constraints affect the scope of biological variation and dynamic patterning through formal dependency relations. Reasoning about constraints can help identify the extent to which these patterns are dependent on mechanistic details, and thereby help answer the question about why these arise.

We contend that not only are insights into formal constraints needed for constructing many mechanistic explanations, but also that knowing what *cannot* be physically or biologically instantiated has explanatory autonomy in biology. In order to clarify how such explanations differ from explanatory strategies focused on mechanistic details and context-dependency, we have provided a scheme of constraint-based explanations that highlight features central to several

abstract strategies in biology, including design explanations and topological explanations. We have further pointed to other kinds of non-mechanistic this class may also contain. In particular, we stress that what distinguishes these from mechanistic explanations is their emphasis on formal constraints and generalizable dependency-relations.

We have provided arguments and examples of why we believe that formal constraints are not only guides or secondary components to mechanistic explanations but also have an explanatory autonomy of their own. Strategies for developing constraint-based explanations are not a threat to mechanistic accounts, except for those that hold that all explanations in biology are mechanistic and context-specific. While mechanistic explanations explain how dynamic patterns are causally realized, constraint-based elucidate how different mechanisms relate to a larger class by outlining dependencies between organization and behavior. Constraint-based explanations can answer questions related to the boundaries of causal explanations – to what is logically, physically and biologically *possible* – whereas mechanistic explanation address questions about how actual causal effects are produced in time. Due to these complementary affordances, constraint-based explanations should not be judged by the same standards for explanatory power as mechanistic explanations.

As we have seen in the case of robust perfect adaptation, the two types of explanation inform each other without being reducible to one another. Constraint-based explanations do not aim at explaining specific systems but at clarifying why the same higher-level patterns appear in different contexts. One can, therefore, interpret some mathematical abstractions as definitions of general relations accounting for a class of systems operating under similar formal constraints. Because of the generality of these constraints, constraint-based explanations have a unifying power typically not associated with biological explanations. Nonetheless, unification should not be taken to imply reductive inferences of the specific from the general. Rather, constraint-based explanations operate at a different level of abstraction, where the same formalization may apply to many different systems with heterogeneous causal detail. Constraint-based explanations thus reflect a concern with identifying the situations where changes to lower-level details do not affect system-level effects.

Our account of constraint-based explanation emphasizes that causal explanations of specific systems are not always the aim of biological research. Although constraint-based explanations can play an important role as part of dynamical mechanistic explanations, we find that classifying all explanations as mechanistic would underestimate the diversity of explanations in biology. Distinguishing between the two types of explanations allows us to conceptualize sources of disagreements due to differences in explanatory standards among different groups of researchers, and understanding the innovative potential of bringing together insights from different explanatory projects.*

* We thank Melinda Fagan, Robert Batterman, Ingo Brigandt, Maria Serban, William Bechtel, Raphael Scholl and several anonymous reviewers for helpful comments and suggestions to earlier version of this paper. We also thank the Center for Philosophy of Science at the University of Pittsburgh for financial support during the initial stages of this paper.

References

- Alon, U. 2007, *An introduction to systems biology: Design principles of biological circuits*. Boca Raton: Chapman and Hall
- Barkai, N., & Leibler, S. 1997, "Robustness in simple biochemical networks", *Nature*, 387(6636), pp. 913-917
- Batterman, R. W. 2010, "On the Explanatory Role of Mathematics in Empirical Science," *British Journal for the Philosophy of Science* 61, pp. 1-25
- Batterman, R. W. 2002, *The devil in the details: Asymptotic reasoning in explanation, reduction, and emergence*, Oxford: Oxford University Press
- Batterman, R. W., & Rice, C. C. 2014, "Minimal model explanations", *Philosophy of Science* 81(3), pp. 349-376
- Bechtel, W., & Abrahamsen, A. 2011, "Complex biological mechanisms: Cyclic, oscillatory, and autonomous", in: C. Hooker (Ed.), *Philosophy of complex systems* (pp. 257-285), Amsterdam: Elsevier
- Bechtel, W., & Abrahamsen, A. 2012, "Thinking dynamically about biological mechanisms: Networks of coupled oscillators", *Foundations of Science* 18, pp. 707-723
- Bechtel, W., & Richardson, R. C. 1993, *Discovering complexity: Decomposition and localization as strategies in scientific research*, Princeton, New Jersey: Princeton University Press.
- Berger, R. 1998, "Understanding science: Why causes are not enough". *Philosophy of Science* 65(2), pp. 306-332
- Bertalanffy, L. v. 1969, *General systems theory. foundations, development, applications*, New York: George Braziller
- Braillard, P.-A. 2010, "Systems biology and the mechanistic framework", *History and Philosophy of the Life Sciences* 32, pp. 43-62.
- Braillard, P.-A. & Malaterre, C. 2015 (Eds.), *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*, New York: Springer
- Brigandt, I. 2013, "Systems biology and the integration of mechanistic explanation and mathematical explanation", *Studies in History and Philosophy of Biological and Biomedical Sciences* 44(4), pp. 477-492.
- Chen, C., Cui, J., Zhang, W., & Shen, P. 2007, Robustness analysis identifies the plausible model of the bcl-2 apoptotic switch, *FEBS Letters* 581(26), pp. 5143-5150

Preprint version. The final version is published in *Dialectica*, Sept 2016.
DOI: 10.1111/1746-8361.12145.

- Craver, C. 2007, *Explaining the brain, mechanisms and the mosaic unity of neuroscience*, Oxford: Clarendon Press
- Csete, M. E., & Doyle, J. C. 2002, "Reverse engineering of biological complexity", *Science* 295(5560), pp. 1664-1669
- Darden, L. 2002, "Strategies for discovering mechanisms: Schema instantiation, modular subassembly, forward/backward chaining", *Philosophy of Science* 3, pp. 354-365
- Eldar, A., Dorfman, R., Weiss, D., Ashe, H., Shilo, B., & Barkai, N. 2002, "Robustness of the BMP morphogen gradient in drosophila embryonic patterning", *Nature* 419(6904), pp. 304-308
- Fagan, M. 2012, "Waddington redux: Models and explanation in stem cell and systems biology", *Biology and Philosophy* 27(2), pp. 179-2013
- Fagan, M. B. 2015, "Stem cells and systems models: clashing views of explanation", *Synthese* 193(3), pp. 873-907.
- Green, S., Fagan, M., & Jaeger, J. 2015, "Explanatory integration challenges in evolutionary systems biology", *Biological Theory* 10(1), pp. 18-35.
- Green, S., Levy, A., & Bechtel, W. 2015, "Design sans adaptation", *European Journal for Philosophy of Science* 5(1), pp. 15-29.
- Green, S. 2015, "Revisiting generality in the life sciences: Systems biology and the quest for general principles", *Biology and Philosophy* 30, pp. 629-652.
- Gross, F. 2015, "The Relevance of Irrelevance: Explanation in Systems Biology", in: C. Malaterre, & P. Braillard (Eds.), *Explanation in Biology* (pp. 175-198). Springer Netherlands
- Hamilton, A. 2007, "Laws of biology, laws of nature: Problems and (dis) solutions", *Philosophy Compass* 2(3), 592-610
- Hesse, M. B. 1963, *Models and analogies in science*, London and New York: Sheed and Ward.
- Hornberg et al. 2005, "Principles behind the multifarious control of signal transduction", *FEBS*, 272:244-258
- Humphreys, P. 2004, *Extending ourselves: Computational science, empiricism, and scientific method*, New York: Oxford University Press
- Huneman, P. 2010, "Topological explanations and robustness in biological sciences", *Synthese* 117, pp. 213-245
- Iglesias, P. A. 2013, "Systems biology: The role of engineering in the reverse engineering of biological signaling", *Cells* 2(2), pp. 393-413.

- Jones, N. (2014), "Bowtie structures, pathway diagrams, and topological explanation". *Erkenntnis* 79(5), pp. 1135-1155
- Jones, N. 2016, "Inference to the more robust explanation". *British Journal for the Philosophy of Science*.
Published online: doi:10.1093/bjps/axw009
- Jones, N., & Wolkenhauer, O. 2012, "Diagrams as locality aids for explanation and model construction in cell biology", *Biology & Philosophy* 27(5), pp. 705-721
- Kauffman, S. 1969, "Metabolic stability and epigenesis in randomly constructed genetic nets", *Journal of Theoretical Biology* 22(3), pp. 437-467
- Kitano, H. 2007, "Towards a theory of biological robustness", *Molecular Systems Biology* 3, p. 137
- Kitano, H. 2004, "Biological robustness", *Nature Reviews Genetics* 5(11), pp. 826-837
- Kitano, H., & Oda, K. 2006, "Robustness trade-offs and host-microbial symbiosis in the immune system", *Molecular Systems Biology*, article number: 2006.0022
- Lange, M. 2015, "Explanation, Existence and Natural Properties in Mathematics—A Case Study: Desargues' Theorem", *dialectica* 69.4, pp. 435-472.
- Lenski, R. E., Barrick, J. E., & Ofria, C. 2006, "Balancing robustness and evolvability". *PLoS Biology* 4(12), e428
- Levy, A., & Bechtel, W. 2013, "Abstraction and the organization of mechanisms", *Philosophy of Science* 80, pp. 241-261
- Matthiessen, D. 2015, "Mechanistic explanation in systems biology: Cellular networks", *The British Journal for the Philosophy of Science*, published online: doi:10.1093/bjps/axv011
- Morange, M. 2005, *Les secret du vivant: Contra la pensée unique en biologie*, Paris: La Découverte
- Morange, M. 2011, "Recent opportunities for an increasing role for physical explanations in biology", *Studies in History and Philosophy of Biological and Biomedical Research* 42, pp. 139-144
- Nersessian, N. J. 2008, *Creating scientific concepts*. Cambridge: The MIT Press
- Orth, J. D., Thiele, I., & Palsson, B. Ø. 2010, "What is flux balance analysis?", *Nature Biotechnology* 28(3), pp. 245-248
- Oster, G., & Desoer, C. 1971, "Tellegen's theorem and thermodynamic inequalities", *Journal of Theoretical Biology* 32(2), pp. 219-241
- Pincock, C. 2011, *Mathematics and Scientific Representation*, New York: Oxford University Press
- Ross, L. 2015, "Dynamical models and explanation in neuroscience", *Philosophy of Science* 82(1), pp. 32-54

Preprint version. The final version is published in *Dialectica*, Sept 2016.
DOI: 10.1111/1746-8361.12145.

Saikumar, P., Dong, Z., Mikhailov, V., Denton, M., Weinberg, J. M., & Venkatachalam, M. A. 1999, "Apoptosis: Definition, mechanisms, and relevance to disease", *The American Journal of Medicine* 107(5), pp. 489-506

Shinar, G., & Feinberg, M. 2011, "Design principles for robust biochemical reaction networks: What works, what cannot work, and what might almost work", *Mathematical Biosciences* 231(1), pp. 39-48

Simon, H. 1977, *Models of discovery*. Dordrecht: D. Reidel

Sober, E. 1983, "Equilibrium explanation", *Philosophical Studies* (43), pp. 201-210

Thompson, D. 1917/2004, *On growth and form*, Cambridge: Cambridge University Press

Tokar, T., & Ulicny, J. 2012, "Computational study of the mechanism of the bcl-2 apoptotic switch", *Physica A: Statistical Mechanics and its Applications* 391(23), pp. 6212-6225.

Tyson, J. 1991, "Modeling the cell division cycle: Cdc2 and cyclin oscillations", *Proceedings of the National Academy of Sciences of the United States of America* 88(16), pp. 7328-7332.

Wolkenhauer, O., Shibata, D., & Mesarović, M. 2012, "The role of theorem proving in systems biology", *Journal of Theoretical Biology* 7(300), pp. 57-61.

Wouters, A. 2007, "Design explanation: Determining the constraints on what can be alive", *Erkenntnis* 67, pp. 65-80.

Yi, T. M., Huang, Y., Simon, M. I., & Doyle, J. 2000, "Robust perfect adaptation in bacterial chemotaxis through integral feedback control", *Proceedings of the National Academy of Sciences of the United States of America* 97(9), pp. 4649-4653.

Youle, R. J. 2007, "Cell biology. Cellular demolition and the rules of engagement", *Science* 315(5813), pp. 776-777.