Diagrams as locality aids for explanation and model construction in cell biology

Nicholaos Jones · Olaf Wolkenhauer

Abstract   Using as case studies two early diagrams that represent mechanisms of the cell division cycle, we aim to extend prior philosophical analyses of the roles of diagrams in scientific reasoning, and specifically their role in biological reasoning. The diagrams we discuss are, in practice, integral and indispensible elements of reasoning from experimental data about the cell division cycle to mathematical models of the cycle's molecular mechanisms. In accordance with prior analyses, the diagrams provide functional explanations of the cell cycle and facilitate the construction of mathematical models of the cell cycle. But, extending beyond those analyses, we show how diagrams facilitate the construction of mathematical models, and we argue that the diagrams permit nomological explanations of the cell cycle. We further argue that what makes diagrams integral and indispensible for explanation and model construction is their nature as locality aids: they group together information that is to be used together in a way that sentential representations do not.

Keywords   Cell biology · Diagram · Explanation · Locality aid · Model construction

N. Jones (✉)
Department of Philosophy, University of Alabama in Huntsville, Morton Hall 332, Huntsville, AL 35899, USA
e-mail: nick.jones@uah.edu

O. Wolkenhauer
Department of Systems Biology and Bioinformatics, University of Rostock, Rostock, Germany

O. Wolkenhauer
Stellenbosch Institute for Advanced Study (STIAS), Wallenberg Research Centre, Stellenbosch University, Stellenbosch, South Africa

Introductory remarks

Cells multiply through periodic cycles of division. Each division cycle has four phases. Sufficient cell growth during G1 phase triggers S (synthetic) phase, during which a cell synthesizes new copies of its DNA. After a short G2 phase, where the cell grows and duplicates much of its "hardware" and checks that DNA replication is complete, the cell enters M (mitotic) phase, during which the cell's old and new DNA separate and the cell divides to form two separate daughter cells, each with a copy of the entire genetic material and other cellular machinery. After division, each daughter cell occupies the G1 phase, and the cell cycle is complete.

In 1991, Tyson and Goldbeter separately published papers that proposed two mathematical models of the cell cycle. Both models purport to explain why the cell cycle oscillates. At the time, this phenomenon was something of a mystery, because no obvious oscillating signal triggers the G1 phase. Tyson and Goldbeter's models show how cyclin-Cdc2 interactions alone suffice for this behavior, because solutions for the differential equations in their models exhibit sustained oscillations. Our interest in these models arises due to the role of diagrams in their construction. This process proceeds in two stages: an initial stage, in which one constructs a diagram using available evidence and some educated guessing; and a final stage, in which one constructs a set of differential equations from the diagram.

Diagrams like Tyson and Goldbeter's allow biologists to represent, archive, exchange, integrate, and reuse knowledge about biological networks obtained through scientific experimentation and data analysis (Le Novère et al. 2009; Saraiya et al. 2005). Most interdisciplinary research projects in biology (and especially systems biology) provide examples in which diagrammatic reasoning plays a central role, not only in the communication between "modeler" and "experimentalist" but also in the formulation and testing of hypotheses. According to Perini, "a thorough philosophical analysis of science will have to include an understanding of how visual representations contribute to the articulation and defense of scientific claims" (2005b: 283). However, the philosophical literature largely overlooks the roles of diagrams in scientific reasoning.

Brown (1997), Perini (2005a, b), Bechtel and Abrahamsen (2005), and Goodwin (2010) are exceptions. Brown argues that diagrams (he calls them pictures) can facilitate the discovery of mathematical theorems and provide defeasible evidence for their truth; but he does not provide much detail about *how* they can do so. Perini (2005b) argues that diagrams are capable of supporting conclusions, and being supported by other representations, because they can be true or false. But, like Brown, she does not discuss *how* diagrams make such contributions. She also argues that the two-dimensional nature of diagrams allows them to provide functional explanations in an especially concise manner and, in some cases, to make that information humanly comprehensible (Perini 2005a). Bechtel and Abrahamsen describe some advantages diagrams offer over sentential representations (2005: 427–428). They sketch an explanation for these advantages based upon the fact that "information that may be only implicit in linguistic representation may be made explicit, and hence easier to invoke in reasoning, in a diagram" (Bechtel and Abrahamsen 2005: 429). However, they do not identify what it is about diagrams that gives them these features, and they

restrict their account to mechanistic explanations (similar to what Perini calls functional explanations).[1] Finally, Goodwin argues that diagrams, and specifically structural formulas from organic chemistry, can be the principle means for expressing a scientific theory, and he discusses the way in which the norms for constructing and interpreting structural formulas change through time (2010: 631–633).

Using Tyson and Goldbeter's diagrams of the cell division cycle as case studies, we aim to extend extant philosophical analyses of the roles of diagrams in scientific reasoning, and specifically their role in biological reasoning. Unlike the diagrams that Goodwin discusses, Tyson and Goldbeter's diagrams are not endpoints of reasoning (in the way that a scientific theory is). Instead, they are, in practice, integral and indispensable elements of reasoning from experimental data about the cell division cycle to mathematical models of the cycle's molecular mechanisms. In accordance with Brown, Perini, and Bechtel and Abrahamsen's analyses, the diagrams provide functional explanations of the cell division cycle and facilitate the construction of mathematical models for that cycle.[2] Extending beyond these analyses, we offer an account of *how* diagrams do this: they are locality aids for information search and retrieval, grouping together information that is to be used together in a way that sentential representations do not. We also show how, as locality aids, diagrams facilitate nomological explanations of the cell cycle.

Our discussion begins, in the next section, by presenting Tyson and Goldbeter's diagrams. We discuss the evidential basis for each diagram, as well as the kinds of assumptions both Tyson and Goldbeter make in constructing their diagrams. Then we argue that each diagram provides a functional explanation of the cell division cycle. We argue that Perini's account of why diagrams are especially well-suited to providing these explanations is incomplete; and we offer a way to extend her account by construing diagrams as locality aids. In the subsequent section, we explain the reasoning process that produces mathematical models from each diagram, identifying both the assumptions with which one constructs kinetic equations from the diagrams and the additional assumptions required to solve these equations. In the next two

---

[1] Mechanistic explanations "account for the behavior of a system in terms of the functions performed by its parts and the interactions between these parts" by identifying the "parts and their organization, showing how the behavior of the machine is a consequence of the parts and their organization" (Bechtel and Richardson 2010: 17). Functional explanations involve "analyzing a disposition d of a [system] a into a number of other dispositions $d_1 \ldots d_n$, had by a or components of a such that programmed manifestations of the d_i results in or amounts to a manifestation of d," where a programmed manifestation is one that "could be specified in a program or a flow chart" (Cummins 1975: 759). Piccinini and Craver (forthcoming) maintain that functional explanations are incomplete mechanistic explanations by virtue of omitting structural aspects regarding the location, shape, orientation, and organization of a system's components; Bechtel and Richardson, that elements of mechanistic explanations include elements of functional ones (2010: 89–90).

[2] These models are, in turn, the basis for formulating and testing hypotheses about the biological mechanisms that regulate the cell cycle. The diagrams also can be used to summarize published data. Through peer review, these summaries can be considered to be firmly established. However, in most cases the models involve uncertainty, so that the diagrams can be considered to be encodings of hypotheses. In such cases, the mathematical models also encode hypotheses, and computational simulation experiments with the model can be brought into agreement (or disagreement) with experimental data in order to validate the hypotheses through the diagram, in combination with mathematical modeling and computational simulation. Further discussing these issues would take us beyond the scope of this paper.

sections, we argue that Tyson and Goldbeter's diagrams are integral and indispensable elements of this reasoning process by virtue of being locality aids for information extraction, and that being locality aids also makes the diagrams especially well-suited for providing nomological explanations. These results further extend Perini's analysis, which addresses neither reasoning with diagrams nor their role in nomological explanation; they extend Bechtel and Abrahamsen's analysis, which only considers the role of diagrams in mechanistic (or functional) explanations; and they extend Brown's analysis, which does not address how diagrams facilitate the construction of mathematical models. They also highlight the usefulness of the notion of a locality aid for explaining the role of diagrams in biological reasoning.

From experimental data to diagrams

Research to identify the molecular mechanisms regulating the cell division cycle began no earlier than the 1960s, as scientists developed genetic and molecular biological methods to study the cell's constituents. By the early 1990s, these investigations revealed two kinds of molecule that regulate the cell cycle, cyclins and cyclin-dependent kinases (CDKs; see Nurse 2000: 75). Cyclins are proteins that periodically (cyclically) change their level of concentration during the cell cycle. These cyclins bind to CDKs, which maintain a constant concentration throughout the cycle but vary their activity depending upon whether they are attached to cyclin molecules. (The cyclin-dependent kinases involved in human cell division cycles are called Cdc2, for cell division control 2, or CDK1, for cyclin-dependent kinase 1).

By 1991, molecular and genetic experiments suggested that a protein complex, maturation (or mitosis) promoting factor (MPF), largely controls the mitotic (M) phase of the cell division cycle. When cyclin, newly synthesized from amino acids, combines with pre-existing Cdc2, the molecules together form the hetero-dimer MPF. A phosphate group then attaches to the cyclin subunit of this MPF complex, and then a different phosphate group detaches, in an autocatalytic way, from the Cdc2 subunit. This dephosphorylation converts the MPF complex from an inactive form to an active one. If a protein kinase does not oppose this dephosphorylation (if, that is, the phosphate group detaches successfully without replacement), active MPF then stimulates several processes essential for nuclear and cellular division, following which the MPF complex dissociates into its components and the cyclin subunit quickly degrades. When a phosphate group attaches to the surviving Cdc2 subunit, the cycle repeats itself–unless the phosphorylation reverses itself. (This summary condenses the discussion in Tyson 1991. See Schafer 1998 and Nurse 2000 for more comprehensive reviews.)

Tyson (1991) constructs a diagram of the cell cycle by consolidating and simplifying the preceding experimental knowledge. (He cites at least 10 papers as the basis for his verbal summary of the overall mitotic cycle, which has been largely reproduced above; see 1991: 7328–7329.) His summary description ignores phosphorylation of other proteins involved in the mitotic cycle, and it ignores the way in which active MPF stimulates cyclin degradation. The summary also attributes cyclin degradation entirely to dephosphorylation, even though
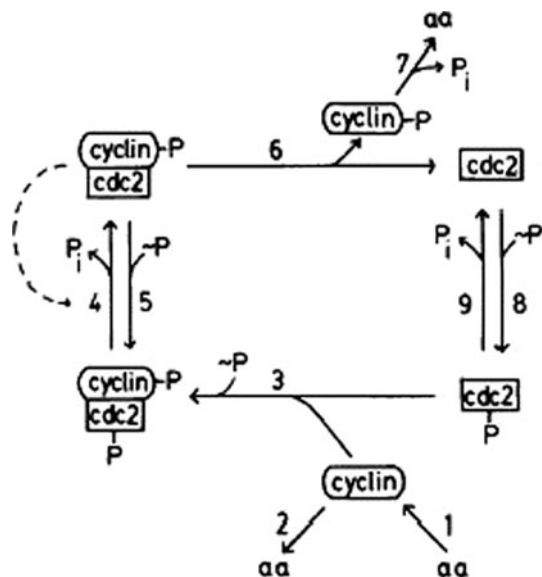
then-available experimental evidence suggested that degradation also requires conjugation with the protein ubiquitin (Tyson 1991: 7329). But, given the basis of evidence he does not simplify away, Tyson produces Fig. 1.

The diagram in Fig. 1 represents nine distinct potential "steps" of the cell division cycle. In Step 1, amino acids (aa) synthesize new cyclin. If the cyclin is unstable, it degrades (Step 2); if it is stable, it combines with a phosphorylated Cdc2 unit (bottom right) to form a cyclin-Cdc2-P complex (Step 3). After a phosphate group (P) combines with the cyclin subunit of this complex (bottom left), the phosphate group on the Cdc2 subunit detaches (Step 4). (The dashed arrow represents autocatalytic feedback of active MPF (P-cyclin-Cdc2 heterodimers) on its own production.) If this process does not reverse itself (Step 5), the complex becomes active MPF (top right). Active MPF dissociates into its cyclin-P and Cdc2 subunits (Step 6). The cyclin-P unit degrades into a phosphate group and amino acids (Step 7), while the Cdc2 unit (top right) undergoes phosphorylation into Cdc2-P (Step 8). This Cdc2-P unit (bottom right) is then available to combine with cyclin, unless the phosphate group detaches (Step 9).

Goldbeter (1991) constructs a diagram of the cell division cycle that represents less biochemical detail than Tyson's. While the purpose of Tyson's paper is to further understand mechanistic details of cell-cycle regulation (1991: 7329), the purpose of Goldbeter's is

> to show, by means of a simple theoretical model, how thresholds in Cdc2 kinase activation and in cyclin degradation may naturally arise as a result of post-translational modification, and how the mitotic cascade involving cyclin and Cdc2 kinase can oscillate as a result of both the time delays associated with these thresholds and the triggering by Cdc2 kinase of rapid cyclin degradation (1991: 9107).

Fig. 1 Tyson's diagram of cell division cycle (1991: 7238)

Experimental evidence available at the time showed that cyclin accumulation promotes the activation of Cdc2, and that active Cdc2 promotes cyclin degradation. Goldbeter takes this to suggest that the oscillatory behavior of the cell division cycle might be a result of a negative feedback loop, and he aims to confirm this possibility by constructing a model for mitotic oscillations that includes a negative feedback loop (1991: 9110).

Goldbeter assumes that cyclin is synthesized at a constant rate, that phosphorylation inactivates Cdc2 and dephosphorylation activates it, that cyclin drives Cdc2 activation, that activated Cdc2 elicits the activation of an unknown "protease" that degrades cyclin, and that cyclin can degrade spontaneously (1991: 9107–9108). These assumptions, all of which have some degree of experimental support, together with the conjecture that there is a negative feedback loop in the cell cycle, yield the diagram in Fig. 2.

Cyclin synthesis occurs at rate v_i, and its spontaneous degradation occurs at a rate v_d. The dashed arrow beginning at the "cyclin" symbol represents cyclin triggering the transformation of inactive Cdc2 (M+) into active Cdc2 (M). This transformation occurs at rate v_1, while the reverse transformation occurs at rate v_2. The dashed arrow beginning at the "M" symbol represents active Cdc2 triggering the transformation of an inactive protease (X+) into an active form (X). This occurs at a maximum rate v_3; the reverse transformation, at a maximum rate v_4. Finally, the dashed arrow beginning at the "X" represents the active protease degrading cyclin. The diagram includes a negative feedback loop, because while cyclin synthesis distally stimulates activation of X, activation of X stimulates cyclin degradation.

Diagrams as locality aids for functional explanation

In accordance with Perini's (2005a) analysis, both Tyson and Goldbeter's diagrams provide functional explanations of the cell division cycle, in the sense of Cummins (1975 see especially 760–761). Tyson's diagram represents the relevant parts of the

Fig. 2 Goldbeter's diagram of cell division cycle (1991: 9108)
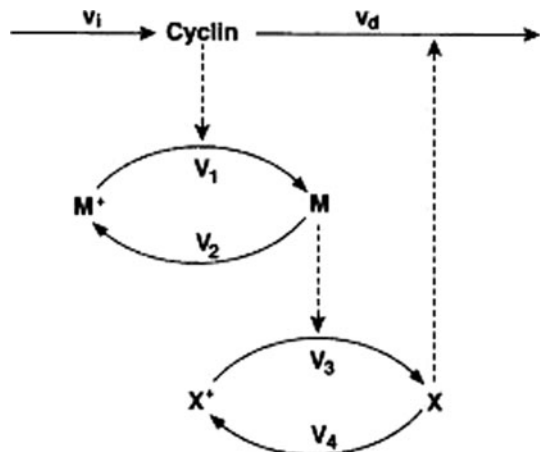
cycle (amino acids, cyclin, Cdc2, phosphate groups), their organization and interaction (combined or not, interacting or not), and their capacities (to activate, deactivate, or combine into other parts); the numbered arrows represent the way in which the cycle is a consequence of the organization and capacities of the parts. Goldbeter's diagram ignores the details by which activation and deactivation of both M and X occur, in order to "avoid entering into the detailed description of a process which is not yet completely clarified" (Goldbeter 1991: 9107) and, perhaps, in order to develop a simple representation. For simplicity, it also ignores the combination of cyclin and Cdc2 into a heterodimer complex, possible modifications of M, differences among kinds of cyclin involved in activating Cdc2, and the nature of the protease X (1991: 9108). But, despite this higher degree of abstraction, Goldbeter's diagram represents relevant parts of the cell division cycle (cyclin, Cdc2, an unknown protease) and the capacities of each part (to activate or deactivate some other part); and the arrows represent the way in which the cycle is a consequence of the parts and their capacities.

Perini (2005a) argues that the visible form of diagrams explains why diagrams are especially well suited to provide functional explanations. Specifically, she argues that their two-dimensional nature allows diagrams to represent simultaneously both the components of a system and their relations to each other, and that this ability of diagrams makes them well-suited for functional explanation because it makes them more concise than their sentential counterparts. The main support for her thesis is a diagram of the binding change mechanism for ATP production in which "the arced shapes … refer to enzyme subunits, and contiguity relations among the symbols for subunits refer to the relation of being in the same complex" (2005a: 266). The diagram's two-dimensional nature allows it to represent simultaneously enzyme subunits and the relations among these subunits. According to Perini, this representational simultaneity makes diagrams more concise, because "in order to describe all the different kinds of relations involved in the model, a series of statements would be required, or a long conjunction," and "the visible form of such a linguistic representation bears no relation to the structure of the model" (2005a: 266).

Perini's account applies nicely to Tyson's diagram. There are symbols that represent molecular components; their proximity to each other at the "corners" of Tyson's diagram concisely represents synchronic binding relations among those components; and arrows between each group of component symbols represent transformations of molecular compounds. Tyson's diagram exhibits this kind of representational simultaneity because it is two dimensional, and it is more concise than its sentential counterpart because it simultaneously represents molecular components and (both synchronic and diachronic) relations among those components.

However, Perini's account fits less well with Goldbeter's diagram. While Goldbeter's diagram represents simultaneously molecular components (with letters) and diachronic relations among those components (with arrows), this does not make his diagram more concise than a sentential representation of the same system. We can construct a sentential representation such that there is a one-to-one mapping between its constituent symbols and the symbols in Goldbeter's diagram. First, introduce symbols to represent molecular components and reaction rates; these can

be the same symbols that occur in Goldbeter's diagram contains ("Cyclin," "M+," "v_i," and so on). Next, introduce a series of predicate letters to represent diachronic relations among these components. For example, "S(x,y)" might represent "x is synthesized at rate y," "D(x,y)" might represent "x degrades at rate y," "T(x,y,z,w)" might represent "x triggers the transformation of y into z at rate w," and so on. After introducing a predicate letter for each arrow in Goldbeter's diagram, construct the sentential representation: S(Cyclin, v_i); D(Cyclin, v_d); T(Cyclin, M+, M, v_1); and so on. While this sentential representation uses abbreviations, Goldbeter's does too. Since the abbreviated sentential representation is not clearly *longer in length* than Goldbeter's diagram, it conveys at least the same amount of information in the same amount of space.

Because equality with respect to brevity entails equality with respect to conciseness, the ability of Goldbeter's diagram to represent simultaneously system components and relations among those components does not make his diagram more concise than its sentential counterpart. Hence, Goldbeter's account cannot be especially well suited to provide a functional explanation of the cell cycle *because it is more concise than its sentential counterpart*. Perini's account, therefore, offers only a partial explanation of why diagrams are especially well suited for functional explanation. A more complete account, which applies to both Tyson and Goldbeter's diagrams, involves the ability of diagrams to be locality aids.

A representation is a *locality aid* when it groups together information that is to be used together, allowing users to minimize the amount of search required to extract from the representation information required to make appropriate inferences (see Koedinger 1992: 151–152). Koedinger gives, as an example, a high school geometry problem (Fig. 3).

Prior to the tenth step in the solution in Fig. 3, there are nine statements that might contribute to making a triangle congruence inference; but a triangle congruence rule (such as the side-angle-side or side–side-side rules) requires only three statements. There are, accordingly, 84 possible statement combinations to
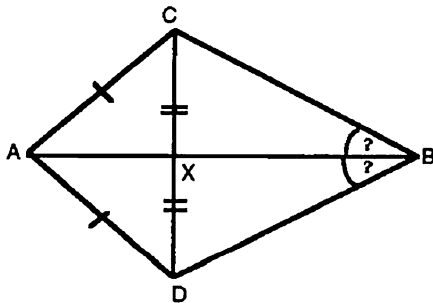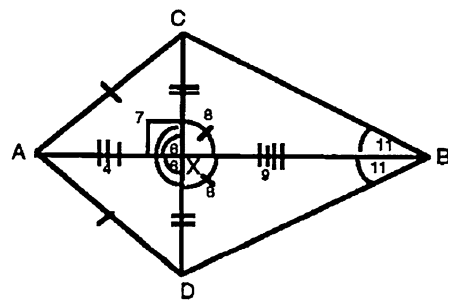
Fig. 3 Geometry problem and solution—sentential form (Koedinger 1992: 152)

Problem statement in sentential form:
  "Given $\overline{AC} \cong \overline{AD}$ and $\overline{AXB}$ bisects $\overline{CXD}$, prove that $\angle CBX \cong \angle DBX$."

Solution in sentential form:

Statements:	Reasons:
1. $\overline{AC} \cong \overline{AD}$	1. Given
2. $\overline{AXB}$ bisects $\overline{CXD}$	2. Given
3. $\overline{CX} \cong \overline{DX}$	3. Def-bisector (2)
4. $\overline{AX} \cong \overline{AX}$	4. Reflexive
5. $\triangle ACX \cong \triangle ADX$	5. SSS (1 3 4)
6. $\angle AXC \cong \angle AXD$	6. Corres-Parts (5)
7. $\overline{AB} \perp \overline{CD}$	7. Eq-Linear-Angs (5)
8. $\angle BXC \cong \angle BXD$	8. Eq-Linear-Angs (7)
9. $\overline{BX} \cong \overline{BX}$	9. Reflexive
10. $\triangle BCX \cong \triangle BDX$	10. SAS (3 8 9)
11. $\angle CBX \cong \angle DBX$	11. Corres-parts (10)

a Problem in diagrammatic form:

b Solution in diagrammatic form:

Fig. 4 Geometry problem and solution—diagrammatic form (Koedinger 1992: 153)

consider when attempting to infer that triangles BCX and BDX are congruent. The sentential solution does not group the relevant statements together; but the diagrammatic solution does, as shown in Fig. 4.

The diagram in Fig. 4 groups together the information that appears as steps 3, 8, and 9 in the sentential solution. This grouping minimizes the amount of search required to find the information that would support inferring triangle congruence with the side-angle-side rule.

Tyson and Goldbeter's diagrams of the cell division cycle are well suited for providing functional explanations of the cell division cycle by virtue of being locality aids, grouping together information that is required for such explanation. Tyson's diagram groups together information about the molecular components of the cell division cycle and their relations to each other, representing one molecule binding to another by having the symbols for molecules touch each other. For example, Tyson's diagram represents cyclin-Cdc2 binding by having the "cyclin" and "Cdc2" symbols share a common border. Goldbeter's diagram does not do this. But it does group together information about the capacities of each molecule, representing the products of these capacities with outgoing arrows. For example, the dashed arrow from the symbol "cyclin" to the arrow labeled "v_1" represents the capacity of cyclin to transform inactive Cdc2 (M+) into active Cdc2 (M) at rate v_1; and the arrow labeled "v_d," departing from "cyclin" to the right, represents an additional capacity of cyclin to degrade spontaneously. The diagram groups these capacities together, because the two arrows depart from the same symbol, "cyclin." A sentential representation of these capacities also could group this information together; but the cost would be ungrouping the representations of dual capacities for other molecules, such as the protease X. Tyson's diagram likewise groups together information about the multiple capacities of different molecules. Moreover, Goldbeter's diagram groups together information about capacities in a way that makes apparent the role of the posited negative feedback loop; and Tyson's diagram groups together information about molecular capacities in a way that makes apparent the cyclical nature of the molecular interactions.

This analysis suggests that, while Perini is correct to note that the two-dimensional nature of diagrams allows them to represent information relevant to functional explanations, her account of the connection between two-dimensionality

and suitability for functional explanation is incomplete. Having two dimensions allows diagrams to be locality aids in a way that their sentential counterparts are not, even when (as in Goldbeter's case) the diagrams are not more concise than their sentential counterparts. For while the ability of diagrams to represent simultaneously system components and relations among those components *sometimes* makes diagrams more concise, sometimes it allows them to be locality aids despite *not* being more concise. Accordingly, the way to remedy the limitation of Perini's analysis is to add that diagrams are especially well suited to provide functional explanations when they are more concise than their sentential counterparts *or when they are locality aids for functional explanation*.

From diagram to mathematical model

The case that the notion of a locality aid helps to explain the role of diagrams in biological reasoning does not rest entirely upon the argument that the sentential counterpart to Goldbeter's diagram is at least as concise as the diagram itself. The notion of a locality aid also accounts for how diagrams facilitate the construction of mathematical models and provide other, non-functional kinds of explanations. Tyson and Goldbeter's diagrams of cell division cycle mechanisms are the means by which they construct mathematical models of that mechanism. These constructions follow what Kell and Knowles refer to as a classical modeling strategy: construction of a structural model representing system elements and reaction relations among those elements, derivation of ordinary differential equations from the model, and subsequent parameterization of these equations (2010: 13). Kell and Knowles do not discuss the details of the transition from structural models (what we here refer to as diagrams) to ordinary differential equations. We shall, accordingly, provide details for this strategy in the case of modeling cell cycle oscillations. This extends Perini's (2005a) and Brown's (2005) analyses, by showing how diagrams facilitate the construction of mathematical models. In the subsequent sections, we use these details to argue that, by virtue of being locality aids, diagrams are especially well-suited for constructing mathematical models and providing non-functional explanations.
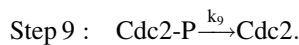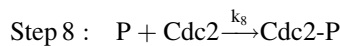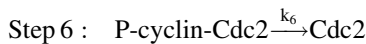
Tyson and Goldbeter's constructions of mathematical equations from their diagrams, and the subsequent biophysical interpretation of the terms in those equations, rely upon mass action kinetics. Specifically, Tyson and Goldbeter apply a (phenomenological) mass-action rate law to the reaction relations represented in their diagrams. These rate laws determine kinetic equations for the changes in concentrations of various biochemical species with respect to time. For example, according to the law of mass action, "the rate of any given elementary reaction is proportional to the concentrations of the species reacting in the elementary process (reactants)" (Crampin et al. 2004: 80). For the elementary reaction in which A irreversibly transforms into B with velocity v:

$$A \xrightarrow{v} B,$$

the law of mass action entails that the rate of change of the concentration of B is proportional to the concentration of B. The proportionality constant (or rate

constant) for such a reaction is normalized by the reaction's stoichiometry (the difference between the amount of the species present after the reaction and the amount of the same species present before the reaction). While the proportionality constant depends upon environmental conditions such as the temperature at which the reaction occurs, applications of mass-action rate laws typically assume that these conditions are invariant (Crampin et al. 2004: 81; Millat et al. 2007: 43).

We mention the law of mass action only to illustrate the idea of a mass-action rate law. The reactions that appear in Tyson and Goldbeter's diagrams are more complex than our toy example, by virtue of being composed of several elementary reactions. A more general mass-action rate law, based upon the idea that biochemical reactions are decomposable into smaller independent steps, applies to more complex reactions, according to which "the rate of change of the concentration of any given species is … a sum of the rates of change due to the elementary reactions in which that species participates" (Crampin et al. 2004: 81; for a more sophisticated generalization, see Heinrich and Schuster 1996: 14–16). Tyson uses such a generalized mass-action rate law (hereafter MARL) to construct several kinetic equations from his diagram of the cell cycle mechanism (see Sible and Tyson 2007: 240). For example, the top right corner of his diagram (Fig. 1) indicates that Cdc2 participates in three elementary reactions:
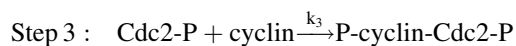
$$\text{Step 6}: \quad \text{P-cyclin-Cdc2} \xrightarrow{k_6} \text{Cdc2}$$

$$\text{Step 8}: \quad \text{P} + \text{Cdc2} \xrightarrow{k_8} \text{Cdc2-P}$$

$$\text{Step 9}: \quad \text{Cdc2-P} \xrightarrow{k_9} \text{Cdc2}.$$

The stoichiometry for Step 6 and Step 9 is (1–0) = 1; for Step 8, it is (0–1) = −1. Applying a generalized MARL to these elementary reaction relations, it follows that the overall change in Cdc2 concentration with respect to time is

$$d[\text{Cdc2}]/dt = k_6[\text{P-cyclin-Cdc2}] - k_8[\text{P}][\text{Cdc2}] + k_9[\text{Cdc2-P}],$$

where a species name in square brackets represents the concentration of that species and k_i is the proportionality constant for the ith elementary reaction step. This is precisely the equation that Tyson constructs, and a similar procedure yields all but two of his equations (see Tyson 1991: 7329, Table 1). (For thoroughness, we note that, at Step 3, Tyson assumes that intermediate reactions in which a phosphate group attaches to a cyclin operate at quasi-steady states. This allows him to ignore these reactions. We shall discuss the quasi-steady-state approximation in connection with Goldbeter's constructions.)

The kinetic equations for which Tyson departs from this relatively straightforward procedure concern rate changes in the concentrations of P-cyclin-Cdc2-P and P-cyclin-Cdc2. For example, the lower left corner of Tyson's diagram (Fig. 1) indicates that P-cyclin-Cdc2-P participates in three elementary reactions:

$$\text{Step 3}: \quad \text{Cdc2-P} + \text{cyclin} \xrightarrow{k_3} \text{P-cyclin-Cdc2-P}$$

$$\text{Step 4a}: \quad \text{P-cyclin-Cdc2-P} \xrightarrow{k_{4'}} \text{P-cyclin-Cdc2}$$

$$\text{Step 4b}: \quad \text{P-cyclin-Cdc2} \xrightarrow{k_4} \text{P-cyclin-Cdc2-P}$$

$$\text{Step 5}: \quad \text{P} + \text{P-cyclin-Cdc2} \xrightarrow{k_5} \text{P-cyclin-Cdc2-P}.$$

The autocatalytic reaction in Step 4 consists of a forward reaction with rate constant $k_{4'}$ when there is no active MPF (P-cyclin-Cdc2) and a reverse reaction (the auto-catalytic feedback loop) with rate constant k_4 when the total concentration of active MPF is equal to the total concentration of Cdc2. (A conservation law determines the value of the total Cdc2 concentration.) According to mass action kinetics, the rate of the entire autocatalytic reaction in Step 4 is

$$F = k_{4'} + k_4([\text{P-cyclin-Cdc2}]/[\text{total Cdc2}])^2.$$

Given this caveat, a generalized MARL entails that

$$\begin{aligned}
d[\text{P-cyclin-Cdc2-P}]/dt = &\; k_3[\text{Cdc2-P}][\text{cyclin}] - F[\text{P-cyclin-Cdc2-P}] \\
&+ k_5[\text{P}][\text{P-cyclin-Cdc2}].
\end{aligned}$$

A similar procedure reproduces Tyson's equation for $d[\text{P-cyclin-Cdc2}]/dt$.

Numerically solving Tyson's equations requires assigning values to the rate constants and the initial concentrations of each biochemical species. Tyson simplifies this task in four ways. First, he assumes that the concentrations of amino acids and phosphate are constant, thereby reducing his set of kinetic equations from nine to seven. Second, he assumes that the total concentration of Cdc2 is constant and equal to the combined concentrations of Cdc2, Cdc2-P, P-cyclin-Cdc2-P, and P-cyclin-Cdc2, thereby further reducing his set of equations by one. Third, Tyson assumes that newly synthesized cyclin is stable, so that $k_2 = 0$, and he ignores rephosphorylation of the Cdc2 subunit of active MPF in Step 5, so that $k_5 = 0$. He notes, however, that his solutions are fairly robust when these parameters have nonzero values. Finally, he assumes that Cdc2 is phorphorylated immediately after it dissociates from P-cyclin-Cdc2, so that $k_8[\text{P}]$ is much greater than k_9, which in turn is much greater than k_6. This entails that $d[\text{Cdc2}]/dt$ is approximately equal to $-k_8[\text{P}][\text{Cdc2}]$; Tyson takes it to warrant treating the Cdc2 concentration as constant.

Tyson further assumes that k_4 is much greater than $k_{4'}$, presumably because it is natural to treat the rate at which active MPF stimulates its own production as much faster than the rate at which active MPF is produced initially. He then assigns values to the other rate constants and, in the case of k_6, considers a case in which the value of the "constant" varies in time. (As noted, rate constants generally are not constant, depending as they do upon environmental conditions.) This parameterization is largely speculative due to lack of experimental data (Tyson 1991: 7329). Accordingly, while the solutions to Tyson's parameterized equations exhibit oscillatory behavior, the uncertainty of the parameterization translates into Tyson characterizing his mathematical model as merely a possible explanation of how the cell cycle mechanism generates oscillations.

Like Tyson, Goldbeter constructs a set of kinetic equations from his diagram using a (phenomenological) mass-action rate law. However, Goldbeter's diagram of the cell cycle mechanism is less firmly grounded in biochemical evidence than

Tyson's, and Goldbeter explains the cell cycle's oscillatory behavior as caused by negative feedback rather than auocatalytic processes (Goldbeter 1991: 9111). The basis for his explanation is Michaelis–Menten kinetics. These kinetics apply to reactions in which a protein substrate S and enzyme E reversibly combine into an enzyme-substrate complex C, which then irreversibly creates a modified form of the original protein, S^*, and the original enzyme E:
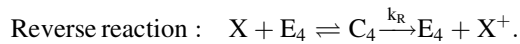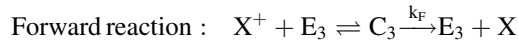
$$S + E \rightleftharpoons C \xrightarrow{k_2} E + S^*,$$
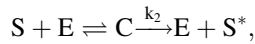
where the forward reaction from $S + E$ to C occurs at rate k_1 and the reverse reaction from C to $S + E$ occurs at rate k_{-1}. Assuming that the total enzyme concentration $[E^T] = [E] + [C]$ is constant (conservation law) and that the intermediate complex forms and dissociates back into its components is much faster than it converts into its products (quasi-steady-state approximation), a generalized MARL entails that the rate at which the concentration of the modified protein S^* changes is

$$d[S^*]/dt = k_2[E^T][S]/(k_M + [S]),$$

where the Michaelis constant k_M is defined as $(k_{-1}+k_2)/k_1$. This rate law is known as the Michaelis–Menten equation; it is valid when either k_M is very large or the substrate concentration is much greater than the enzyme concentration. (For a fuller discussion of the validity conditions for the Michaelis–Menten equation and the relation between this equation and a generalized MARL, see Millat et al. 2007: 41–44.) When the intermediate complex concentration in a reversible Michaelis–Menten reaction is negative compared to the concentration of the protein in forms S and S^*, similar assumptions entail that the rate at which the concentration of S^* changes is a sum of two Michaelis–Menten-like rates, where one rate characterizes the change in the concentration of S^* in the forward reaction and the other characterizes the change in the concentration of S^* in the reverse reaction (see Heinrich and Schuster 1996: 17–18).

Goldbeter uses a generalized MARL, together with some simplifying auxiliary assumptions, to construct three kinetic equations from his diagram of the cell cycle mechanism. The most important auxiliary assumptions, which allow him to apply the Michaelis–Menten equation, are the quasi-steady-state approximation and two conservation assumptions to the effect that the total concentrations of the kinase M and the protease X (in their active and inactive forms) remain constant. Goldbeter makes several further assumptions as well. But rather than explicate his reasoning in its entirely, we offer a partial reconstruction of one equation in order to highlight to role of Goldbeter's diagram in the construction process.

The solid arrows at the bottom of Goldbeter's diagram (Fig. 2) indicate that the active (X) form of the protease participates in two elementary reactions. Goldbeter's diagram does not explicitly represent these reactions as Michaelis–Menten type, nor does it indicate the enzymes involved in the reactions; but his subsequent discussion, and accepted convention, support constructing these reaction relations as follows:

$$\text{Forward reaction}: \quad X^+ + E_3 \rightleftharpoons C_3 \xrightarrow{k_F} E_3 + X$$

$$\text{Reverse reaction}: \quad X + E_4 \rightleftharpoons C_4 \xrightarrow{k_R} E_4 + X^+.$$

(The solid arrow departing from the symbol "X" indicates that X promotes cyclin degradation.) Applying the Michaelis–Menten equation to these reaction relations entails that the overall rate of change in the concentration of the active protease X is

$$d[X]/dt = k_F [E_3^T][X^+]/(K_{MF} + [X^+]) - k_R [E_4^T][X]/(K_{MR} + [X]),$$

where K_{MF} and K_{MR} are, respectively, the Michaelis constants for the Forward Reaction and the Reverse Reaction. Goldbeter assumes that $[E_3^T]$ is proportional to [M], which entails that $k_F[E_3^T] = k_{F'}[E_3^T][M]$. A few further manipulations yield the precise equation Goldbeter constructs, and similar procedures yield his other two equations (see Goldbeter 1991: 9108).

Like Tyson, Goldbeter parameterizes his mathematical model in order to show that solutions to his equations exhibit oscillatory behavior. This involves some degree of speculation. But it does not interfere with his stated goal. Goldbeter is very careful to claim that his model demonstrates that a negative feedback loop is a *potential* cause of the cell cycle's oscillations: in Goldbeter's words, "[the model's] analysis highlights the conditions in which the cyclin-cdc2 kinase system can operate as a continuous autonomous oscillator" (1991: 9107). Goldbeter achieves this goal, because solutions to his parameterized equations exhibit oscillatory behavior without oscillating input (since his model assumes a constant rate of cyclin production). And it does so by virtue of applying a generalized mass-action rate law to the reaction relations constructed from his diagram.

Diagrams as locality aids for model construction

Both Tyson and Goldbeter use diagrams of potential cell cycle mechanisms to construct a set of kinetic equations for those mechanisms. The diagrams are integral to these constructions, because they are locality aids. The diagrams group together, in an especially efficient way, information about reaction relations between components of the proposed cell cycle mechanisms; and this grouping minimizes the amount of search required to extract information relevant to the construction of mathematical models.

There are two reasons to suppose that Tyson and Goldbeter's diagrams are locality aids for constructing mathematical models of the cell division cycle. First, they group together information from a variety of publications that report experimental data about the cell division cycle. For example, Tyson cites Nurse (1990) as part of his diagram's experimental basis. That paper reports that activation of MPF determines whether the cell enters M phase, and that this activation requires dephosphorylation of Cdc2-P and association with cyclin. Tyson's diagram synthesizes many such reports, ignoring much of their content and representing only those biochemical details (in the form of nine reaction relations) that suffice for

a "first approximation" of the overall mechanism. Specifically, Tyson constructs his diagram from six sources (1991: 7329); but by 1991 there are orders of magnitude more papers reporting experimental results related to the cell division cycle. Goldbeter's diagram similarly consolidates experimental data from many publications.

Second, Tyson and Goldbeter's diagrams group together information about the cell cycle in a way that makes minimizes the amount of search required to extract that information for model construction. For example, Tyson's diagram makes evident that calculating the rate of change for the concentration of P-cyclin-Cdc2-P requires taking into account exactly three reaction relations, because the diagram shows that there are exactly three arrows associated with the P-cyclin-Cdc2-P icon. The diagram also makes evident what those relations are. Tyson's sentential description of his diagram's content, in contrast, requires searching through multiple lines of text to find the number and nature of the relevant reactions. Tyson's diagram facilitates the construction of nine reaction relations from experimental data; and while the sentential description of the diagram's content facilitates the same construction process, that process likely would require more significant indirect processing, in the form of searching through multiple lines of text, were the diagram absent.

Diagrams as locality aids for nomological explanation

In addition to making Tyson and Goldbeter's diagrams especially well-suited for constructing mathematical models, being locality aids for information extraction makes their diagrams especially well-suited for nomological explanation. Following Hempel (1965), a nomological explanation of a phenomenon is a derivation of that phenomenon from law statements and auxiliary assumptions. Tyson and Goldbeter both deduce equations that exhibit oscillatory behavior, which is the phenomenon they seek to explain. These deductions begin with the construction of reaction relations from their respective diagrams. From these relations, Tyson and Goldbeter derive a set of differential equations using a (phenomenological) mass-action rate law and auxiliary assumptions. After making further assumptions that parameterize these equations, Tyson and Goldbeter deduce solutions to their equations that exhibit oscillatory behavior. By subsuming this behavior under laws in such a manner, they provide nomological explanations of why the cell division cycle oscillates.

The mass-action rate laws are essential ingredients for Tyson and Goldbeter's nomological explanations of oscillation of the cell division cycle. So, too, are the reaction relations. Tyson and Goldbeter construct these relations from their respective diagrams. While, in principle, they could construct the relations entirely from sentential representations of the cell division cycle, in practice the diagrams are integral components of their construction (see also Sible and Tyson 2007: 239–240). The diagrams facilitate the constructions by grouping together information about the cell cycle in a way that minimizes the amount of search required to extract information relevant to constructing the reaction relations and deriving reaction rate equations. For this reason, being locality aids for constructing

mathematical models of the cell cycle makes the diagrams especially well-suited to providing material for nomological explanations of why the cell division cycle oscillates.

Concluding remarks

Brown argues that diagrams are integral elements of reasoning in mathematics because they "provide the *known to be true* consequences that we use for testing" hypotheses (1997: 165). According to Brown, diagrams enlarge our pool of intuitive truths and thereby provide a kind of inductive evidence for the truth of mathematical theorems (1997: 167–169, 177). Tyson and Goldbeter's diagrams are not like this. While they are syntheses of available experimental data, they are not therefore evidence that can be used to confirm or refute mathematical models.[3] This is, in part, because they can contain conjectures not warranted by available evidence (as in Goldbeter's case), and in part because their use involves a tacit assumption that no other factors are present. Our sources of evidence regarding mathematical models of biological phenomena are the phenomena themselves rather than diagrams that purport to represent those phenomena.

Despite this difference between mathematical and biological diagrams, diagrams are integral elements of reasoning in biology. For they facilitate the construction of mathematical models of biological phenomena, and they are especially well suited for representing information relevant to functional and nomological explanations. The aim of this paper has been to exhibit *how* diagrams in biology fulfill these roles, and to offer an explanation of *why* diagrams are integral and indispensable tools in fulfilling these roles despite the availability of sentential representations with equivalent content.

We agree with Perini that diagrams are superior to sentential representations because they represent simultaneously both components of a system and their relations to each other. Perini accounts for this superiority in terms of conciseness. But Goldbeter's diagram of the cell division cycle gives some reason to suppose that diagrams are superior even when they are not more concise. Identifying diagrams as locality aids overcomes this (potential) limitation of Perini's analysis. While any diagram that is a locality aid simultaneously represents both components of a system and their relations to each other, not all diagrams that simultaneously represent in this way are more concise than their sentential counterparts. Goldbeter's diagram is a case in point.

The notion of a locality aid explains how, as Bechtel and Abrahamsen note, diagrams can make explicit information that remains implicit in sentential representations. By grouping together information that is to be used together, diagrams make explicit relationships the identification of which would require extended search within sentential representations. The extended discussion of how one constructs mathematical models from Tyson and Goldbeter's diagrams not only illustrates this but also exhibits the role of diagrams in reasoning and nomological

---

[3] But see our previous footnote.

explanation within cell biology. Finally, the fact that being a locality aid is, in a sense, a product of the ability to represent simultaneously both system components and relations among those components helps to explain why the visual format of diagrams makes them especially well suited for providing functional explanations.

References

Bechtel W, Abrahamsen A (2005) Explanation: a mechanistic alternative. Stud History Philos Biol Biomed Sci 36:421–441

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

Brown JR (1997) Proofs and pictures. Br J Philos Sci 48:161–180

Crampin EJ, Schnell S, McCharry PE (2004) Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. Prog Biophys Mol Biol 86:77–112

Cummins R (1975) Functional analysis. J Philos 72:741–765

Goldbeter A (1991) A minimal cascade model for the mitotic oscillator involving cyclin and Cdc2 kinase. Proc Natl Acad Sci 88:9107–9111

Goodwin W (2010) How do structural formulas embody the theory of organic chemistry? Br J Philos Sci 61:621–633

Heinrich R, Schuster S (1996) The regulation of cellular systems. Chapman & Hall, New York

Hempel CG (1965) Aspects of scientific explanation and other essays in the philosophy of science. Free Press, New York

Kell DB, Knowles J (2010) The role of modelling in systems biology. In: Szallasi Z, Stelling J, Periwal V (eds) System modeling in cellular biology: from concepts to nuts and bolts. MIT Press, Cambridge, MA, pp 3–18

Koedinger KR (1992) Emergent properties and structural constraints: advantages of diagrammatic representations for reasoning and learning. AAAI technical report SS-92-02, pp 151–156

Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, Bergman FT, Gauges R, Ghazal P, Kawaji H, Li L, Matsuoka Y, Villéger A, Boyd SE, Calzone L, Courtot M, Dogrusoz U, Freeman TC, Funahashi A, Ghosh S, Jouraku A, Kim S, Kolpakov F, Luna A, Sahle S, Schmidt E, Watterson S, Wu G, Goryanin I, Kell DB, Sander C, Sauro H, Snoep JL, Kohn K, Kitano H (2009) The systems biology graphical notation. Nat Biotechnol 27:735–741

Millat T, Bullinger E, Rohwer J, Wolkenhauer O (2007) Approximations and their consequences for dynamic modeling of signal transduction pathways. Math Biosci 207:40–57

Nurse P (1990) Universal control mechanism regulating onset of M-phase. Nature 344:503–508

Nurse P (2000) A long twentieth century of the cell cycle and beyond. Cell 100:71–78

Perini L (2005a) Explanation in two dimensions: diagrams and biological explanation. Biol Philos 20:257–269

Perini L (2005b) The truth in pictures. Philos Sci 72:262–285

Piccinini G, Craver C (forthcoming) Integrating psychology and neuroscience: functional analyses as mechanism sketches. Synthese. doi:10.1007/s11229-011-9898-4

Saraiya P, North C, Duca K (2005) Visualizing biological pathways: requirements analysis, systems evaluation and research agenda. Inform Vis 4:1–15

Schafer KA (1998) The cell cycle: a review. Vet Pathol 35:461–478

Sible JC, Tyson JJ (2007) Mathematical modeling as a tool for investigating cell cycle control networks. Methods 41:238–247

Tyson JJ (1991) Modeling the cell division cycle: Cdc2 and cyclin interactions. Proc Natl Acad Sci 88:7328–7332