Synthetic Biology and Synthetic Knowledge

Christophe Malaterre
Assistant professor of philosophy of science
Philosophy department
Université du Québec à Montréal,
Montréal, QC, Canada
malaterre.christophe@uqam.ca

Abstract

Probably the most distinctive feature of synthetic biology is its being "synthetic" in some sense or another. For some, synthesis plays a unique role in the production of knowledge that is most distinct from that played by analysis: it is claimed to deliver knowledge that would otherwise not be attained. In this contribution, my aim is to explore how synthetic biology delivers knowledge via synthesis, and to assess the extent to which this knowledge is distinctly synthetic. On the basis of distinctions between knowledge-how and knowledge-why, and between syntheses that succeed and syntheses that fail, I argue that the contribution of synthesis to knowledge is best understood when syntheses are construed as experimental interventions that aim at probing causal relationships between properties of the entities that are combined through these syntheses and properties of their target products. The distinctiveness of synthetic biology in its quest for knowledge through synthesis stems from its ability to sample at will a space of empirical possibilities that is not only huge but also that has been so scarcely sampled by nature.

Keywords

synthetic biology, synthetic knowledge, knowledge-how, knowledge-why, making as knowledge, synthesis, analysis

Today's synthetic biology bears little resemblance to that of Stéphane Leduc at the start of the 20th century (Keller 2002): it somehow encompasses, redefines, and broadens the field of biotechnology (Koide et al. 2009). Some of its ultimate goals include the design and construction of complete genetic and biomolecular systems attached to specific organisms in order to make them capable of reading specific signals, processing them, and producing desired outputs. For some, "synthetic biology is the engineering of biology: the synthesis of complex, biologically based (or inspired) systems, which display functions that do not exist in nature" (Serrano 2007: 1). As its very name suggests, "synthesis" is at the core of

synthetic biology: synthetic biology is a biological discipline for which making, assembling, constructing – in a word: synthesizing – biological entities is essential. For some, this distinctive trademark of synthetic biology leads to an epistemic specificity and specific forms of knowledge. It has been argued that "synthesis drives discovery and paradigm changes in ways that analysis cannot" (Benner et al. 2011: 88), and that it leads to a specific form of knowledge, namely "making as knowledge" (O'Malley et al. 2007, Keller 2009). Such claims rely on two assumptions: first the assumption that synthetic biology focuses on synthesis as research methodology and as such behaves differently from the rest of biology, which is viewed as relying on analysis; second the assumption that synthesis as such produces knowledge that would otherwise not be attained. In this contribution, my aim is to explore how synthetic biology generates knowledge via synthesis, and to assess the extent to which this knowledge is distinctive of a synthetic form of research. To this aim, I first review how synthesis is appealed to in the field of synthetic biology. I then explicate the notion of synthesis – as opposed to analysis – in the context of synthetic biology research. For the sake of analyzing the type of distinctive knowledge – if any – produced by synthetic biology, I then distinguish between two different types of knowledge, knowledge-how and knowledge-why. I then analyze the type of knowledge that is produced by a synthesis depending on whether this synthesis is perceived as a success or a failure. I argue that a successful synthesis delivers both knowledge-how and knowledge-why, the latter being best construed as causal knowledge linking properties of the entities that are combined through the synthesis and properties of the entities that result from the synthesis. In particular, I propose to construe syntheses as specific interventions within causal models that aims at explaining properties of the entities that result from these syntheses. Furthermore, I argue that the distinctiveness of synthetic biology in its quest for knowledge through synthesis stems from its ability to sample at will a space of empirical possibilities that is not only huge but also that has been so scarcely sampled by nature. In the case of failure, I argue that one must distinguish between syntheses that fail to deliver their target products and which lead to little, if any, knowledge, and syntheses that deliver their end-products, yet of properties of which are not those that were initially expected. Contrary to common parlance, I propose that such syntheses be understood as successful and that their contribution to knowledge be analyzed accordingly.

Synthesis and Knowledge in Synthetic Biology

Synthetic biology covers a broad range of bio-engineering activities, somehow extending the field of biotechnology into more radical modifications of living organisms (Koide et al. 2009). There is however no firm consensus about what specifically falls within or outside synthetic biology: for some, synthetic biology is about modifying living organisms with "biobricks"; for others, it is about synthesizing living organisms from scratch including their macromolecules; and for others still, it is about complete genome re-engineering. As a matter of fact, it has been argued that synthetic biology covers at least three broad types of

research activities: (i) the engineering of genetic circuits, (ii) the engineering of entire genomes, and (iii) the engineering of organisms (O'Malley et al. 2007; Malaterre 2009). The design and production of biological oscillatory networks and biochemical switches (e.g. Elowitz and Leibler 2000; Tigges et al. 2009; Kim et al. 2006) typically falls within the first type of synthetic biology, as do the "rewiring" of existing genetic circuits that are made to respond to other molecular signals (e.g. Dueber et al. 2004), and the more systematic usage of "biobricks" as means of implementing given sets of functions within "chassis organisms" (Endy 2005). The second type of research that also receives the name of synthetic biology includes the *de novo* synthesis of whole genomes like those of the smallpox virus (Cello et al. 2002) or of *Mycoplasma genitalium* (Gibson et al. 2008), as well as genome simplification and redesign projects (e.g. Chan et al. 2005). Finally, the third type of synthetic biology includes research that aims at engineering complete novel living systems from scratch, and that focuses, for instance, on self-assembling liposomes coupled with genetic polymers (e.g. Rajamani et al. 2008), on the synthesis of novel genetic systems (e.g. Benner et al. 2011) and more generally on the synthesis of "protocells" (e.g. Noireaux et al. 2005; Rasmussen et al. 2003; Szostak et al. 2001).

Despite being quite heterogeneous in their objectives, these different activities that fall within the scope of synthetic biology all have one thing in common: they focus on *making* novel organisms, rather than on understanding existing ones. They are driven by action-oriented verbs that relate to the concrete realization of bio-related systems: they aim at *synthesizing*, *creating*, *assembling, manipulating, rearranging* things. The "synthetic" feature of synthetic biology sets it aside from the rest of biology that focuses more on analyzing extant organisms, on understanding their features, how they work and why they are there. In a word, synthetic biology is said to focus on synthesis as research strategy (Benner and Sismour 2005). And this feature clearly differentiates it from the rest of biology that is viewed as endorsing analytical approaches. Microbiology, for instance, focuses on analyzing the features of microorganisms, on characterizing their structural elements and identifying their functional aspects, and all of these approaches are typically analytical.

Yet, if synthetic biology is indeed characterized by a synthetic approach, then an interesting question becomes that of assessing the extent to which synthesis as a research strategy leads to the production of specific knowledge, knowledge that would otherwise not be attained. For Benner and colleagues, synthesis drives the discovery of knowledge in ways that analysis cannot (Benner et al. 2011; Benner, this volume). Such statements are not without similarity to those of Berthelot who, at the end of the 19th century, argued about the decisive role of synthesis, but in chemistry. "By limiting ourselves to analysis", Berthelot said, "we would never be able to reach a perfect knowledge of Nature and our mind would not be entirely satisfied" (Berthelot 1860: xv). If this is also the case in synthetic biology now, then it entails that this discipline – thanks to its synthetic approach – leads to knowledge statements that are out of reach of a form of biological research that would be solely based on analytical approaches. Of course, this does not imply that synthetic biology relies only on synthetic approaches to research. As has been noted before

by O'Malley and colleagues, "analytic practices are just as much in the foreground of synthetic biology as are synthesizing strategies", and all the more so as synthetic biology in general relies on much of the knowledge-base delivered by these analytic approaches (O'Malley et al. 2007: 62). Yet it does imply that there would exist knowledge statements whose empirical corroboration would not be possible without the mobilization of synthetic approaches. In other words, synthetic approaches would be necessary to reach certain types of knowledge, without generally being sufficient in this respect. This has led to the coining of the expression "making as knowing" in the context of synthetic biology (Keller 2009). Yet interestingly, Keller argues that this form of knowledge does not apply well to current synthetic biology, which is better construed as technoscience and whose objectives are not to contribute to an understanding of biology but rather to engineer novel organisms. In this respect, she says, synthetic biology considers "making not as knowing, but as an alternative to (or replacement for) knowing" (Keller 2009: 338). Obviously, whether synthetic biology delivers a specific form of making-based knowledge or not hinges on what one takes to be "synthesis" on the one hand, and "knowledge" on the other.

What is "Synthesis"?

If we refer to Lalande's *Vocabulaire technique et critique de la philosophie*, "synthesis" concerns the combination of two or more entities that together form something new. More specifically, synthesis can be defined as the "act of putting together different things that are first given separately, and of uniting them into a whole" (Lalande [1926] 2002: 1091). It is worth distinguishing "abstract synthesis" from "concrete synthesis". In the case of an *abstract synthesis*, the things that are put together and combined are abstract entities such as statements, sets of statements or arguments. One way to perform an abstract synthesis is to combine simple statements into more complex ones. For instance, one may combine "Socrates is old" and "Socrates is a man" into "Socrates is an old man", thereby synthesizing the first two statements into a third one. Another way to construe an abstract synthesis is to define it as the act of combining premises and of deriving a conclusion, of going from some true statements to other true statements entailed by the first ones. In this respect, combining "Socrates is a man" and "All mean are mortal" into "Socrates is mortal" can be understood as another form of abstract synthesis. A third way to perform an abstract synthesis, as proposed by Lalande ([1926] 2002), is to go from detailed statements to more general ones that abstract away from details. This is what happens in the case of historical syntheses, for example, that lead to general views or narratives on the basis of more detailed facts. When the entities that are combined together in the act of synthesis are not abstract but concrete, one speaks of *concrete synthesis*. Concrete synthesis thereby is the operation by which material entities are combined into a material whole. One may assemble mechanical systems from parts, but one may also combine chemical substances into other chemical substances as is the case, for instance, with the synthesis of organic chemicals.

If anything distinguishes synthetic biology from the rest of biology, it is not abstract

synthesis. In fact, abstract synthesis is an activity that pervades all disciplines of biology (and that obviously applies well beyond biology and the natural sciences). The "modern evolutionary synthesis" for instance – that showed, in the 1930s-1940s, that Mendelian genetics was consistent with Darwinian evolution within the framework of population genetics – is but one major example of abstract synthesis in biology. Abstract synthesis is overall a very common activity in biological research. Detailed findings are combined and abstracted away into more general statements. In microbial diversity research – just to take another specific example – elementary findings about cell abundances in subseafloor sediments are combined with statements about mean sedimentation rates and distances from land, so as to deliver more general statements about global subseafloor sedimentary microbial abundances (e.g. Kallmeyer et al. 2012). It will therefore raise no objection to state that it is not abstract synthesis that distinguishes synthetic biology from the rest of biology. Correlatively, this means that the distinguishing type of synthesis that is attributed to synthetic biology is to be understood as *concrete* synthesis.

At this point though, it seems that more needs to be said on concrete synthesis. Consider, as a first classic example of synthesis, Berthelot's chemical synthesis of acetylene thanks to an electric arc between two carbon electrodes in a container filled with hydrogen gas (Berthelot 1860). This synthesis unfolds according to the following chemical reaction: $2C + H_2 \rightarrow C_2H_2$ (in modern chemical notation). As such, it fits the requirements of a concrete synthesis as defined above, and by which material entities are combined into a material whole: the carbon atoms indeed combine with a hydrogen molecule to form the compound entity called acetylene. Yet concrete syntheses can be more complex that this particular example, and also more tricky to characterize. In fact, consider this second example, also taken from the classic repertoire of chemistry: the synthesis of urea[1]. This synthesis can be carried out by combining lead cyanate and ammonia in water so as to form ammonium cyanate according the reaction: $Pb(NCO)_2 + 2NH_3 + 2H_2O \rightarrow Pb(OH)_2 + 2NH_4(NCO)$. Ammonium cyanate then decomposes to ammonia and cyanic acid, the latter producing urea in a nucleophilic addition followed by tautomeric isomerization, according to: $NH_4(NCO) \rightarrow NH_3 + HNCO \leftrightarrow (NH_2)_2CO$. In this case, one does indeed speak of a synthesis, yet such a synthesis does not, strictly speaking, fit the definition of a concrete synthesis given above. It is hard to identify the material entities that would be combined to form a new material whole, and there is no uniting of entities into a larger whole. In urea, one does not find the initial compounds that were put together at start. Rather, there appear to be precursor entities (lead cyanate, ammonia, water) that react with one another, form intermediate compounds (ammonium cyanate) that, in turn, decompose and change configuration to produce the target compound (urea) and some others along the way. More broadly, it would seem that what characterizes numerous concrete syntheses is not so much the combining of material entities into larger wholes, but the production of a target system – possibly associated with waste systems – from a set of precursor systems upon which different activities are carried out that include combining, splitting and rearranging things.

---

1   This synthesis was first carried out by Wöhler (Wöhler 1828). The chemical reactions presented here are slightly different from those initially investigated by Wöhler, but are taken as classic tokens of the set of chemical reactions that are known to lead to the synthesis of urea.

Many of the syntheses of synthetic biology are indeed more complex than the simple combining of material entities into larger wholes. The experimental realization of biological oscillatory networks, for instance, involves the construction of plasmids as DNA vectors that introduce specific DNA sequences into the genome of target *E. coli* bacteria (e.g. Elowitz and Leibler 2000). Similarly, the *de novo* synthesis of whole genomes such as that of *Mycoplasma genitalium* involves the use of intermediate plasmids as DNA vectors and of *S. cerevisiae* yeasts as temporary genome assembly machines (Gibson et al. 2008). We are therefore faced with a dilemma. Either we adopt a narrow construal of "concrete synthesis" as that of uniting parts into a whole – call it "mereological concrete synthesis" (see Figure 1) – and we are led to exclude many chemical and synthetic biological syntheses that, strictly speaking, do not fit this definition. Or we adopt a broader construal of "concrete synthesis" – call it "productive concrete synthesis" – that is not centered on the parts-whole relationship but simply on the productive aspect of synthesis – exhibited, for instance, in the making/producing/creating of a target system from precursor systems – and we are led to accept as concrete syntheses many other activities in biology that do not belong to the discipline of synthetic biology, such as tissue growing or chimeras development. In this case, synthesis would not be a *unique* feature of synthetic biology as it would also be shared by several other biological disciplines or activities. The claim that synthesis is what distinguishes synthetic biology from the rest of biology should thereby probably be best understood as the claim that synthesis is what *typically* distinguishes synthetic biology from *most* of the rest of biology. With this distinction in mind, Benner's claim that "synthesis drives discovery and paradigm changes in ways that analysis cannot" would remain valid, yet with the *proviso* that it does not apply solely to synthetic biology but to all disciplines that involve some form of "productive concrete synthesis".

Figure 1.
The different meanings of "synthesis" and "analysis", and how they respond
to each other (assumed meanings in synthetic biology are highlighted in grey).

Note that in such claims about the role of synthesis in synthetic biology, "synthesis" is often contrasted with "analysis". According to Lalande's *Vocabulaire technique et critique de la philosophie* again, the term "analysis" has two major meanings (Lalande [1926] 2002). The first one is associated with the idea of solving a formal problem. In this case, analysis is the process of establishing a chain of reasoning, typically from a complex proposition whose truth is to be established back to simpler propositions that are taken for granted. As such, this form of analysis is the reverse of synthesis when construed with one of its abstract meanings (see above): whereas synthesis can be understood as a process running deductively from premises to conclusions, analysis can be understood as a process running from conclusions to premises, the first ones being entailed by the others. The second major meaning of analysis is associated with the idea of decomposing something into more elementary entities. In this case, analysis corresponds to the process of breaking an abstract topic or a concrete object into smaller parts, typically with a view to gain a better understanding of it. When analysis is carried out on an abstract topic, such as a statement, it leads to the identification of more elementary entities that are also abstract. It is worth noting, however, that an analysis of a concrete object can be performed either abstractly or concretely (see Figure 1). In the first instance, analysis leads to specific statements about the concrete object, such as statements about its properties, be they functional or relational. In the second, it leads to the identification of more elementary concrete entities such as parts or modules. Consider an *E. coli* bacteria: it can be analyzed in terms of its functional or phenotypic properties (abstract entities), but it can also be analyzed in terms of organelles and other intra-cellular structures (concrete entities). It is this second type of analysis – call it "mereological concrete analysis" – that is antonymic to the "mereological concrete synthesis" mentioned above[2].

The fact that analytic practices are as much in the foreground of synthetic biology as are synthetic ones (O'Malley et al. 2007) can be understood in light of the conceptual clarifications we have just made. In fact, when dealing with concrete entities, synthetic biology relies both on the concept of "abstract analysis of concrete entities", for instance as a means of identifying properties of sub-cellular systems or of individual macro-molecules, and on the concept of "mereological concrete analysis", for instance as a means of checking the presence of some key structural elements inside some of the entities that have been synthesized. Strictly speaking therefore, "analysis" in synthetic biology is not exactly antonymic to "synthesis": of course, the mereological part of both concepts is, yet both concepts also expand beyond this mereological part.

With these distinctions in mind, Benner's claim that "synthesis drives discovery and

---

2 Of course, mereological concrete analysis can be applied to systems produced through productive concrete synthesis generally speaking. Yet by so doing, one does not always reverse the process by which the productive concrete synthesis has produced its target systems. This may be the case with mereological systems. Yet, this is not so with many other systems produced through productive concrete synthesis. For instance in the chemical case of urea, one may decompose this compound into its atomic constituents through mereological concrete analysis; yet one does not find back the compounds that were initially mixed together. It is for this reason that "mereological concrete analysis" is antonymic to "mereological concrete synthesis", but not to the broader concept of "productive concrete synthesis".

paradigm changes in ways that analysis cannot" is best understood as the claim that "productive-concrete-synthesis drives discovery and paradigm changes in ways that abstract-analysis-of-concrete-entities and mereological-concrete-analysis cannot", and with the *proviso* that this claim does not apply solely to synthetic biology but to all disciplines that rely somehow more heavily on productive concrete synthesis than on abstract analysis of concrete entities and on mereological concrete analysis. The central question that should be addressed can now be rephrased as: Which knowledge does productive concrete synthesis generate in synthetic biology that is not generated elsewhere in biology through abstract analysis of concrete entities or mereological concrete analysis?[3]

Distinctions in Knowledge

There are different types of knowledge that I wish to point out in relationship to the specific question that interests us here. I do not want to enter into the complex debates that are amply tackled in epistemology about the necessary and sufficient conditions for knowledge, the sources of knowledge or its structure and limits. Assuming that knowledge is possible and well characterized – for instance as a form of justified true belief – I will limit my aim to the characterization of two major *types* of knowledge that I think make particular sense in the context of synthetic biology: "knowledge-how" and "knowledge-why". These distinctions draw, among others, from arguments made by Ryle and Polanyi on knowledge-how (Polanyi 1958; Ryle 1949), and arguments formulated by Hintikka about knowledge why, when, where and what (Hintikka 1975). The idea here is to be able to account, on the one hand, for the knowledge that is relevant for experimenting or intervening on concrete objects – and in particular those that synthetic biology focuses on – and on the other hand, for the knowledge one acquires as a result of this experimenting and that concerns the explanation of why these very objects have the properties they have or behave the way they do. While knowledge-how subsumes the practical type of knowledge that is required for manipulating such concrete objects and is the knowledge that makes possible interventions onto these concrete objects, knowledge-why accounts for the type of knowledge that is about these concrete objects and that is produced as a result of interventions onto them or their parts. In short, one could say knowledge-how is relevant to intervening onto Nature, whereas knowledge-why is relevant to understanding Nature.

The way I have just formulated this distinction in knowledge makes particular sense in an interventionist account of causal explanation (Woodward 2003)[4]. Knowledge-why can be understood as an explanation of why a certain phenomenon takes place, as causal

---

3 In the rest of the paper, I will use synthesis as meaning "productive concrete synthesis" and analysis as meaning either "abstract analysis of concrete entities" or "mereological concrete analysis", unless specified otherwise.

4 In this contribution, I do not do much justice to Woodward's account of causal explanation. My aim is simply to frame my argument in ways compatible with an interventionist account of causation that is particularly relevant to an experimental science as synthetic biology. For more details on such account of causation, see (Woodward 2003).

knowledge that concerns this phenomenon. Simply put, it is the knowledge that certain causes produce certain effects. And through this knowledge, we know why these effects take place. Such causal knowledge can typically be gained through various interventions on a relevant set of possible causes. And in this context, knowledge-how can be understood as the type of knowledge that one mobilizes to perform these very interventions that are required to identify the causes of the effects we are interested in. Consider the phenomenon of molecular recognition (Benner et al. 2011). Some macromolecules like DNA exhibit this property. One may ask: Why? What are the causes of this property? To answer this question, if we know *how* to replace some of the constituents of DNA by others, we can carry out interventions onto these constituents, synthesize the modified genetic polymer and assess whether it also has the property of molecular recognition. And, by looking at the results that different interventions have onto the property of molecular recognition, we can identify the causal relationships between the constituents of DNA and its property of molecular recognition, and thereby understand *why* a genetic polymer such as DNA exhibits this property of molecular recognition (I will detail this example below). In short therefore, knowledge-why is the type of knowledge that corresponds to causal explanations of specific phenomena, whereas knowledge-how is the type of knowledge that makes possible the interventions that reveal the causal connections that underlie the phenomena at stake.

Consider now a schematic productive concrete synthesis of the type encountered in synthetic biology. Such concrete synthesis will start from a set of initial entities – call them components c_i – that it will combine by carrying out a certain number of activities a_i with a view to creating a given target system s (see Figure 2). We can now be more specific about our initial question and rephrase it as:

> Which knowledge-how and knowledge-why does the productive concrete synthesis that is carried out upon the components c_i by performing the activities a_i with a view to creating the target system s deliver?

At this point, we need to consider two cases that are often singled out when it comes to assessing the fruitfulness of synthesis: the cases of success and of failure of the synthesis at stake. Indeed, synthesis is claimed to generate knowledge in ways that analysis cannot both when it succeeds and when it fails (e.g. Benner, this volume).
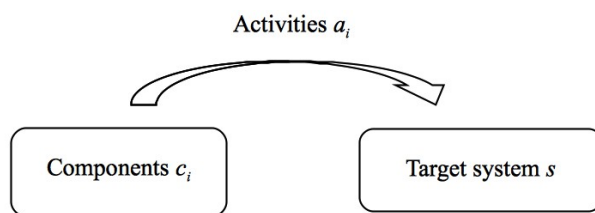

Figure 2.
The core elements of a productive concrete synthesis.

Knowledge from success

The most immediate form of knowledge delivered by a successful productive concrete synthesis is knowledge-how. A successful synthesis tells us indeed that we do know how to produce a given system s from a set of components c_i. This knowledge extends both to the components and activities that are mobilized by the synthesis. When the synthesis is successful, we know that we have well chosen the components c_i, and that, as such, they constitute a sufficient set of components to synthesize s. We also know that we have well chosen the activities a_i that result in the proper combining of the different components c_i into s. We know that these activities are sufficient to bring about system s given the components c_i, and this knowledge typically includes a proper choice of experimental protocol and tools (experimental as well as computational)[5]. This knowledge is all the more patent as it is prone to improvement. As more and more syntheses are carried out with slightly different combinations of components and activities, learning occurs and results, among others, in better success rates, higher quality outputs, lower cost/time ratios, or increased experimental robustness. These improvements may be the result of pure luck, trial and error, or rational decision, yet in any case, they make knowledge-how all the more significant in a synthesis.

Does a successful synthesis also deliver knowledge-why? I take an answer to such a question to be a causal explanation that identifies a set of causes and effects, and their relationships. Following Woodward's account of causal explanation, we can define causation as a relationship between causal variables that is identified through possible interventions onto these variables (Woodward 2003). For instance, given a set of variables **V** and two variables N and R that belong to **V**, one can find out whether N causes R simply by intervening on N and looking at the changes – if any – that are brought about on R, while keeping all the other variables in **V** at some fixed value[6]. The identification of such causal models fits well the experimental approach that prevails in synthetic biology – and in many other experimental disciplines – since it accounts for the practice according to which "by varying one factor, I can make another vary". And it also accounts for the very pragmatic objectives of synthetic biology that is so characteristically goal-oriented in its ambitions to synthesize very specific entities of biological relevance. So, how does synthesis lead to the identification of causal models, and thereby to knowledge-why?

---

5 One may argue that, before answering such questions, one must identify the conditions according to which one knows that a given synthesis is indeed successful. If success is defined as the delivery of the target product s, then the question becomes that of knowing whether s has been delivered or not. It is interesting to note that answering this question requires the application of analytic approaches to the products delivered by the synthesis. It is indeed by analyzing the products, either abstractly (with an identification of its properties) or concretely (with an identification of its parts and structural elements), that one will know whether s is present or not. As mentioned above, analysis and synthesis are much interwoven.

6 A much more precise definition of this interventionist account of causation is found in (Woodward 2003); a cause is defined as being either a "direct" or a "contributing cause", each being defined in reference to the notion of "intervention" through the conditions of "manipulationism" (2003: 59) and of "intervention variable" (2003: 98).

Consider again the example of the causal role of DNA constituents over the property of DNA molecular recognition that Benner and colleagues use in their argument about the knowledge-related role that synthesis plays in synthetic biology (Benner et al. 2011; see also Benner, this volume). As is well-known, DNA is constituted by nucleobases held together by a backbone of ribose sugars and linker phosphates. Molecular recognition consists in the fact that nucleobases exhibit pairing rules that result, among others, in the ability of two DNA strands to align with each other in an antiparallel fashion. These pairing rules consist of molecule-based rules for molecular complementarity: a rule of size-complementarity according to which large nucleobases pair with small ones, and a rule of hydrogen bonding complementarity according to which hydrogen bond donors from one nucleobase pair with hydrogen bond acceptors from the other. By synthesizing an artificial DNA whose naturally occurring nucleobases have been replaced by artificial nucleobases with the same size-bonding and hydrogen-bonding complementarity properties – call this synthesis "synthesis α" – Benner and colleagues were able to create an artificially expanded genetic information system that works in the same fashion that DNA works, thereby corroborating the pairing rules. And by synthesizing artificial DNA whose ribose sugars had been replaced by glycerol – "synthesis β" – and still other DNA whose (electrically charged) phosphate linkers had been replaced by uncharged linkers – "synthesis γ" – Benner and colleagues were able to show that molecular recognition did crucially rely on the presence of ribose sugars and charged phosphate linkers, thereby probing the limits of the nucleobases pairing rules taken in isolation from backbone sugars and linkers. All of these syntheses were successful in so far as they resulted in the desired biomolecules[7]. Yet, to which extent do these syntheses explain the rules of molecular recognition? Consider the simplified causal model in which the effect variable R is molecular recognition and the cause variables N, B, L are respectively the types of nucleobases, of backbones and of linkers (see Figure 3). Assume further that each variable can take two values (R: r_1 = "yes", r_2 = "no"; N: n_1 = "naturally occurring nucleobases", n_2 = "artificial nucleobases"; B: b_1 = "ribose sugar", b_2 = "glycerol"; L: l_1 = "charged phosphate linker"; l_2 = "uncharged linker"). With these notations in mind, synthesis α can be read as a causal intervention onto N (changing $N = n_1$ to $N = n_2$, while holding $B = b_1$ and $L = l_1$). Similarly, synthesis β is an intervention onto B (changing $B = b_1$ to $B = b_2$, while holding $N = n_1$ and $L = l_1$) and synthesis γ is an intervention onto L (changing $L = l_1$ to $L = l_2$, while holding $N = n_1$ and $B = b_1$). Syntheses can therefore be understood as contributing to causal knowledge, and

---

7   Interestingly, Benner and colleagues estimate that, unlike synthesis α which clearly is a success, the two other syntheses β and γ are to be considered failures (Benner et al. 2011). Strictly speaking, I cannot agree as the outcomes of all syntheses were indeed the target systems that the scientists wanted to create: macromolecules of a certain kind composed of very specific entities. In this respect, the syntheses did work according to plan, and do deserve to be called successful. It is true, however, that the outcomes of syntheses β and γ did not exhibit the properties that the scientists had predicted they would (predictions were that changing L or B would not change R). It is therefore only the predictions about the behavior of the target systems that failed, not the syntheses themselves. Once syntheses are understood as interventions, all of them contribute to revealing patterns of functional dependences between causal variables, and these patterns of dependences may be they positively or negatively correlated, or not correlated at all. I address this sense of failure in the following section.

thereby to knowledge-why, in so far as they constitute successful interventions onto causal variables in a causal model. Syntheses help identify the functional relationships that exist between some causal variables that are related to the components c_i that enter the synthesis and some other variables that refer to the target system s itself. They help, for instance, identify the fact that changing the linker component from $L = l_1$ to $L = l_2$ changes the property of molecular recognition of the target genetic polymer from $R = r_1$ to $R = r_2$. They thereby explain the phenomenon of molecular recognition (and its inhibition) by the presence (or the absence) of a particular type of linker component inside the target genetic polymer.
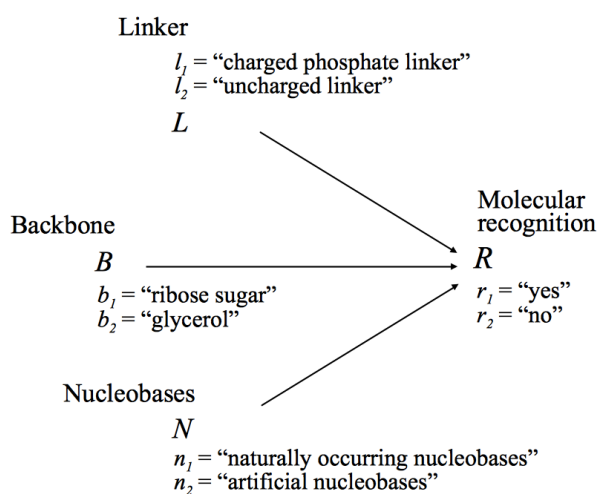
Linker

l_1 = "charged phosphate linker"
l_2 = "uncharged linker"

L

Backbone

Molecular recognition

B

R

b_1 = "ribose sugar"
b_2 = "glycerol"

r_1 = "yes"
r_2 = "no"

Nucleobases

N

n_1 = "naturally occurring nucleobases"
n_2 = "artificial nucleobases"

Figure 3.
Illustration of a causal model with four variables (and their respective value sets) in the case of the synthesis of alternative genetic systems (simplified, based on Benner et al. 2011).

A corollary question is to assess the extent to which synthesis plays a *unique* role in the production of this knowledge. While macromolecules of the kind (l_1, b_1, n_1) occur in nature, the achievement of the synthetic approach is to create macromolecules of different kinds, such as (l_2, b_1, n_1) or (l_1, b_1, n_2), that are not – to our best knowledge – naturally occurring. Of course, one could object that, were we to identify such macromolecules in nature, one would no longer need to rely on synthesis as a means to knowledge. This may be true. And in such cases, one would probably say that it is analysis that has led to the knowledge at stake, insofar as it is analytical approaches that typically lead to the identification of specific entities and their properties in nature. However, synthesis does play a unique role in that it enables one to explore the biochemical space in the direction one deliberately chooses – independently of the identification of naturally occurring entities. In addition, synthesis enables a much thorough and systematic investigation of specific regions of this biochemical space (whereas the exploration of the same biochemical space by means of identifying and analyzing naturally-occurring entities will be much more so fragmented, due to the size of the biochemical space and to the limited sampling that nature can

historically afford[8]). For instance, in the case of genetic molecular recognition, syntheses conducted by Benner and colleagues systematically tackled changes in sugar backbones, in linkers, and in nucleobases. Synthesis therefore appears as an approach that is all the more fruitful and needed as the empirical space of possibilities – one may think here in terms of combinatorial alternatives – is large and scarcely sampled by nature. This is very much so in chemistry, and even more so in biology. Synthesis may therefore be the route to knowledge, and most notably knowledge-why, not for in-principle reasons – since it could ultimately always be replaced by analysis if its target systems were to be found as naturally-occurring ones – but for pragmatic reasons that have to do with the dimension of space of empirical possibilities that is open for us to explore.

Knowledge from failure

Failure is also said to generate knowledge, and sometimes even more so than success (e.g. Benner et al. 2011). In fact, a diagnostic of failure may originate in two very distinct cases. The first – and most obvious – case is when a synthesis does not produce the outcomes that it was expected to produce. We can rephrase this by saying that a synthesis fails when performing the activities a_i upon the components c_i does not result in the target system s. Yet, to which extent can we say that a synthesis that fails in this way delivers knowledge? Consider first knowledge-how. It is doubtful that one would consider "knowing that combining the components c_i by carrying out the activities a_i does *not* produce system s" a proper form of knowledge about "how to make s". Rather, one would say that this is symptomatic of a *lack* of knowledge-how[9]. Of course, this lack of knowledge-how could, in turn, be used as a heuristics. The failure of the synthesis may originate from the set of activities a_i that were performed, hence point to possible changes in experimental protocol and tools. It may also originate from the set of components c_i that were initially selected, and hence point to possible changes in this set of components (and possibly in the set of reasons that had led to this initial choice of components). Note however that a failed synthesis does not uniquely point to a particular element that needs changing. Rather, it is a typical case of epistemic underdetermination (e.g., Duhem 1906; Quine 1951): we know that something is wrong among the set of components c_i and activities a_i that were initially chosen; yet we do not know what precisely is wrong. As such therefore, a failed synthesis is not knowledge but a promise for knowledge. Similarly, consider now knowledge-why as

8   To illustrate this point, the number of DNA strands that are 200 nucleotides long is 4^{200}. This is roughly 10^{120}, hence much more that the 10^{80} particles that the universe is often assumed to include (e.g. Kauffman 2000). Nature is thereby constrained in its exploration of the biochemical space, and so are we. Yet the difference is that nature may have explored parts of this space contingently, whereas we can explore it rationally and at will through synthesis.

9   Strictly speaking however, it could still be argued that since the proposition "combining the components c_i by carrying out the activities a_i does *not* produce system s" is true, it can constitute knowledge. This is in particular the case if knowledge is defined as justified true belief. Yet, in practice, such cases of failure are typically discarded, and it is only cases of success that make it to the records. In this paper, I follow this line of argument.

we have defined it earlier in relationship to characterizing causal relationships. Again, it is doubtful that one would consider a failed synthesis as delivering any form of knowledge-why. Rather, a failed synthesis can be understood as a failed intervention within a causal model. Yet failed interventions are not proper interventions, and as such cannot reveal anything within a causal model.

The second case of failure is when a synthesis does work and produces the target system *s*, yet when one finds out that this system *s* does not have the properties it was expected to have. This is, for instance, what happened with Benner's attempts to synthesize an alternative genetic system by changing the backbone sugar and the linker (Benner et al. 2011). The syntheses as such were successful in so far as they did deliver the target systems, i.e. the modified DNAs. Yet they were also, in some sense, failures in so far as these modified DNAs did not behave as expected, like regular DNA, but somehow folded onto themselves thereby preventing molecular recognition. Strictly speaking therefore, it is not the syntheses that failed, and their contribution to knowledge-how and knowledge-why is best understood as that of successful syntheses (see previous section). Rather, the sense of failure comes from a mismatch between the expected properties of the target systems delivered by the syntheses and their real properties[10]. It is therefore the theories behind the prediction of the properties of the target systems that are disconfirmed in such cases[11]. In other words, it is the causal model that we initially had that proved to be erroneous. In their initial causal model of molecular recognition, Benner and colleagues believed that sugars and linkers had no causal relationship to molecular recognition, the latter being solely a result of molecule-based rules for molecular complementarity, hence a matter of nucleobases. When they – successfully – synthesized alternative DNA strands with a different sugar or a different linker, the resulting macromolecules did not display the property of molecular recognition. This was disappointing for them, hence the sense of failure attributed to these syntheses. Yet, by the same token, these syntheses revealed other patterns of causal relationships that involved causal variables that had initially been discarded. This led Benner and colleagues to revise their causal model of molecular recognition. More generally, I argue that such successful-synthesis-with-a-sense-of-failure are typically those syntheses that point to holes or errors in our causal modeling of the world. Maybe our initial set of causal variables was erroneous: it might have been incomplete, omitting key variables or including irrelevant ones; it might also have assumed wrong value sets for certain variables. Maybe our initial mapping of the causal relationships was erroneous: it might have incorrectly assumed causal relationships between given

---

10 It should be noted that it is not the syntheses *per se* that disconfirm these theories, but the experimental work that is carried out after the syntheses, onto the systems that they have led to. Indeed, it is the subsequent experimental assessments – which also often rely on analytic approaches – that reveal the real properties of the systems that have been successfully synthesized. The achievements of the syntheses are to make possible these further experimental assessments. If the syntheses had not been possible – and if the target systems are not naturally-occurring – then there would be no way to carry out these further experimental assessments, and to expand our knowledge about the properties of these target systems.

11 In this contribution, I take the concept of "theory" rather loosely without committing to any particular view, be it syntactic or semantic. The concept is not central here, and can be understood in a broad range of ways without changing the nature of the argument I am defending.

variables when none exist; it might have omitted others; it might have also assumed incorrect functional relationships between given variables. In such cases, the synthesis raises a flag and indicates that our causal model is wrong somewhere. In turn, because it points to different possible sources of errors of causal modeling, such successful-synthesis-with-a-sense-of-failure plays a crucial heuristic role. Yet, once again, it should be noted that such synthesis never unambiguously points to specific elements that would account for this failure. Rather, it enables us to say that *something* went wrong *somewhere* in our causal modeling of nature. But it does not tell us precisely neither what nor where, as it is still replete with epistemic underdetermination.

Conclusion

"Does synthesis generate knowledge in ways that analysis cannot?". By explicating the different meanings of synthesis and analysis, I have shown that this question is ambiguous and that synthesis and analysis cannot be so directly opposed to each other in synthetic biology. In fact, synthesis and analysis do coexist under different – and not always antonymic – forms in synthetic biology. I have also proposed to distinguish between two different types of knowledge: knowledge-how and knowledge-why. This has led me to reformulate of our initial question as: "Which knowledge-how and knowledge-why does the productive concrete synthesis that is carried out upon the components c_i by performing the activities a_i with a view to creating the target system s deliver?". Distinguishing between cases of success and of failure, I have argued that a successful productive concrete synthesis leads to specific knowledge-how that concerns precisely the components c_i and activities a_i that are required for successfully delivering the target system s. I have further argued that such synthesis also leads to specific knowledge-why that is best understood as being causal knowledge about relevant variables related to the components c_i and to the target system s. No such knowledge is delivered by a synthesis that fails to produce the target system it was intended to produce. While cases of failure are also often taken to include cases of successful syntheses that lead to target systems that do not exhibit the properties that they were expected to, I have argued that the contribution to knowledge of such cases is best understood as that of successful syntheses, and in particular by taking into account their role as interventions within a causal model. The sense of failure that accompanies such syntheses does not come from a failure of these syntheses to produce their target systems – since they do – but arises from the fact that the syntheses point to errors in the causal models we had taken for granted. What appears to be distinctive of synthetic biology's route to knowledge is that it broadens the perimeter of systems that can be manipulated beyond those that are naturally-occurring, and that it does so in an empirical space of biochemical possibilities that is truly large and so scarcely sampled by nature.

Acknowledgements

References

Benner SA, Sismour AM (2005) Synthetic biology. Nature Reviews Genetics 6: 533–543.

Benner S, Chen F, Yang Z (2011) Synthetic biology, tinkering biology, and artificial biology: A perspective from chemistry. In Luisi PL, Chiarabelli C (eds) Chemical Synthetic Biology. Wiley, pp. 69-106

Benner S (this volume) Synthesis as Route to Knowledge. Biological Theory.

Berthelot M (1860) Chimie organique fondée sur la synthèse, Paris: Mallet-Bachelier.

Cello J, Paul A, Wimmer E (2002) Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template. Science 297: 1016–1018.

Chan LY, Kosuri S, Endy D (2005) Refactoring bacteriophage T7. Molecular Systems Biology 1: 0018.

Dueber JE, Yeh BJ, Bhattacharyya RP, Lim WA (2004) Rewiring cell signaling: The logic and plasticity of eukaryotic protein circuitry. Current Opinion in Structural Biology 14: 690–699.

Duhem P (1906) La théorie physique: son objet et sa structure. Paris: Marcel Rivière & Cie.

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcrip- tional regulators. Nature 403: 335–338.

Endy D (2005) Foundations for engineering biology. Nature 438: 449–453

Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden- Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, Glass JI, Venter JC, Hutchison CA III, Smith HO (2008) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science 319: 1215–1220.

Hintikka J (1975) The Intensions of Intentionality and Other New Models for Modalities. Dordrecht. Ch 1.

Kallmeyer J, Pockalnyc R, Adhikari R, Smith D, D'Hondt S (2012) Global distribution of microbial abundance and biomass in subseafloor sediment, Proc Nat Acad Sci USA 109(40): 16213-16216.

Kauffman S (2000) Investigations. New York: Oxford University Press.

Keller EF (2002) Making Sense of Life: Explaining Biological Development with Models, Metaphors, and Machines. Cambridge, MA: Harvard University Press.

Keller EF (2009) Knowing as making, making as knowing: The many lives of synthetic

biology. Biological Theory 4: 333–339

Kim J, White KS, Winfree E (2006) Construction of an in vitro bistable circuit from synthetic transcriptional switches. Molecular Systems Biology: 68

Koide T, Pang WL, Baliga NS (2009) The role of predictive modeling in ra- tionally re-engineering biological systems. Nature Reviews Microbiology 7: 297–305. Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA III, Smith HO, Venter JC (2007) Genome transplantation in bacteria: Changing one species to another. Science 317: 632–638. Lee D, Granja JR, Martinez JA, Severin K, Ghadiri MR (1996) A self-replicating peptide. Nature 382: 525–528.

Lalande A ([1926] 2002) Vocabulaire technique et critique de la philosophie. Paris: PUF

Malaterre C (2009) Can synthetic biology shed light on the origins of life? Biological Theory (4)4: 357-367.

Noireaux V, Bar-Ziv R, Godefroy J, Salman H, Libchaber A (2005) Toward an artificial cell based on gene expression in vesicles. Physical Biology 2: P1–P8.

O'Malley M, Powell A, Davies J, Calvert J (2007) Knowledge-making distinctions in synthetic biology. BioEssays 30: 57-60.

Polanyi M (1958) Personal Knowledge: Towards a Post-Critical Philosophy. Chicago: University of Chicago Press

Quine W (1951) Two Dogmas of Empiricism, Philosophical Review 60: 20-43.

Rajamani S, Vlassov A, Benner S, Coombs A, Olasagasti F, Deamer D (2008) Lipid-assisted synthesis of RNA-like polymers from mononucleotides. Origins of Life and Evolution of Biospheres 38: 57– 74.

Rasmussen S, Chen L, Nilsson M, Abe S (2003) Bridging nonliving and living matter. Artificial Life 9: 269–316.

Ryle G (1949) The Concept of Mind. Hutchinson. Ch 2.

Serrano L (2007) Synthetic biology: Promises and challenges. Molecular Systems Biology 3: 158–162.

Szostak JW, Bartel D, Luisi PL (2001) Synthesizing life. Nature 409: 387–390.

Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M (2009) A tunable synthetic mammalian oscillator. Nature 457: 309–312.

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

Wöhler F (1828) On the artificial production of urea. Annalen der Physik und Chemie, 88.