

Design under Randomness: How Variation Affects the Engineering of Biological Systems

Tero Ijäs

tero.ijas@helsinki.fi

University of Helsinki, Department of Political and Economic Studies, PO Box 24 (Unioninkatu 40), 00014 University of Helsinki, Finland.

Abstract

Synthetic biology offers a powerful method to design and construct biological devices for human purposes. Two prominent design methodologies are currently used. *Rational design* adapts the design methodology of traditional engineering sciences, such as mechanical engineering. *Directed evolution*, in contrast, models its design principles after natural evolution, as it attempts to design and improve systems by guiding them to evolve in a certain direction. Previous work has argued that the primary difference between these two is the way they treat variation: rational design attempts to suppress it, whilst direct evolution utilizes variation. I argue that this contrast is too simplistic, as it fails to distinguish different types of variation and different phases of design in synthetic biology. I outline three types of variation and show how they influence the construction of synthetic biological systems during the design process. Viewing the two design approaches with these more fine-grained distinctions provides a better understanding of the methodological differences and respective benefits of rational design and directed evolution, and clarifies the constraints and choices that the different design approaches must deal with.

Keywords: Synthetic biology; Engineering; Design; Variation; Rational Design; Directed Evolution

1. Introduction

The emerging field of synthetic biology has been framed as a way of bringing an engineering perspective and principles to biotechnology. One of these principles is the higher emphasis on design and, consequently, identifying design as an autonomous phase distinct from fabrication (Cambray et al. 2011; Heinemann and Panke 2006). Successful

separation of design and fabrication should allow higher division of labor between these phases, or even automation of many of the fabrication tasks. Besides emphasis on autonomous design, the influence of engineering principles for synthetic biology is seen in the formulation of systematic design methodologies. The systematization of design methodology is an important feature of most engineering sciences (see Houkes and Vermaas 2010; Kroes 2012), and synthetic biologists have proposed different design approaches to better control the variability and stochastic interaction of biological systems.

I will analyze the role of design approaches in controlling biological complexity from the perspective of biological variation. I focus on two prominent design methodologies: rational design and directed evolution. In Section 2, I provide an analysis of the respective methodological differences between these approaches. *Rational design* is a component-focused approach that resembles the design methodology of traditional engineering sciences, such as mechanical engineering. It is, subsequently, the most prominent design methodology for those synthetic biologists that argue that synthetic biology should model its core methodology after engineering (e.g., Endy 2005; Heinemann and Panke 2006). *Directed evolution*, in contrast, models its design principles after natural evolution, as it attempts to design and improve systems by guiding them to evolve in a certain direction. Despite the prominence of engineering-inspired methods such as rational design, directed evolution or similar evolutionary design approaches are also common in the practice of synthetic biology (e.g., Haseltine and Arnold 2007; Packer and Liu 2015; Romero and Arnold 2009).

These approaches are often juxtaposed both in the biological (e.g., Silver et al. 2014) and philosophical literature (e.g., Krohs and Bedau 2013; Lewens 2013; Morange 2013). These accounts have not, however, dealt with the question of variation or how it relates to the specifics of different design approaches. I will analyze the differences between these design approaches by framing them through their relationship to variation. I follow the categorization offered by Giese et al. (2013) in their bibliometric analysis of synthetic biology. They argue that the primary methodological controversy of synthetic biologists is based on contrary views concerning biological variation and the inherent randomness of biological systems. They distinguish between “rational” and “evolutionary” groups of synthetic biologists.¹ The rational group treats biological variation as problematic and aims to suppress it in its design, whereas the evolutionary group sees variation as a fundamental part of biological systems and aims to utilize it in their design process.

In general, I agree with Giese et al.’s (2013) differentiation between rational and evolutionary groups of synthetic biologists, and their stances toward this variation.

Biological variation and noise can be understood either as a destructive or as a constructive factor in design. However, Giese et al. (2013) are analyzing more general trends in the practice of synthetic biology based on bibliometric analysis. Therefore, whereas their dichotomy is well-suited to analyze differences in the synthetic biology community, it is too coarse to discern actual methodological questions on how variation affects design.

¹ Though Giese et al. (2013) approach the question from the perspective of different groups of synthetic biologists, their categorization of rational and evolutionary groups is equivalent to the distinction between rational design and directed evolution design approaches.

Therefore, in Section 3, I propose a new classification for distinct types of variation in synthetic biology and show how they influence the design process both in rational design and directed evolution. I argue that from the perspective of design practice, the categorization offered by Giese et al. (2013) between variation-suppressing rational design and variation-utilizing evolutionary design is too simplistic, and the specific type of variation must also be considered. I argue that biological variation causes distinctly different problems on the genotypic and phenotypic level for the design of synthetic systems, and analyze how these problems relate to each design approach. I distinguish three types of variation that are relevant to design tasks in synthetic biology: genetic variation (Section 3.1.), functional phenotypic variation (Section 3.2.), and environmental phenotypic variation (Section 3.3.). I demonstrate how each of these variations affects different design choices, how the suppression of variation is more fruitful in some cases than in others and how different types of variation may be correlated.

2. Design in Synthetic Biology

Most instances of synthetic biology are application-oriented as they aim to construct biological systems and devices for human purposes. These desired functions and systems are often not found in nature, and Dougherty and Arnold (2009, p. 486) argue that successful application of synthetic biology will require “parts that are not biologically relevant, parts that solve human problems and not necessarily problems for the organisms that make them.” This is both the rationale and the main challenge of design in synthetic biology, as synthetic biologists need to plan and construct novel structures that would exhibit the required beneficial functionalities.

One should, however, exercise some caution in drawing analogies between engineering and synthetic biology. For one, engineering as a discipline and a practice it is not uniform in its design methodology. When synthetic biology is framed as a way of bringing engineering perspective to synthetic biology, these proponents of synthetic biology usually rely on an overly narrow conception of engineering. For example, Heinemann and Panke (2009, p. 394-395) use the manufacturing of a new car as an example of the typical characteristics of real engineering, suggesting that “we just have to replace the car with biological cell and have to employ the outlined features of engineering disciplines.” Calcott et al. (2015) criticize this image of engineering, which focuses on the design of mechanical and electric machines, as narrow and misleading. They argue that actual engineering is much more heterogeneous in its design methodology and practices; subfields such as software engineering rely on design principles that differ radically from those in mechanical engineering (see also Calcott 2014). Therefore, when proponents of synthetic biology argue that one should “start thinking of designing biological parts in a forward engineering manner” (Heinemann and Panke 2009, p. 392) they should be understood in the context of “classical” engineering, such as mechanical and electrical engineering.

One feature of classical engineering, which synthetic biologists aim to adapt, is the idea that you can distinguish between the design of the artifact and the execution of the design blueprint, that is, fabrication of the artifact. In previous biotechnology and molecular biology, design work was tied to *wet lab* work with no clear division between design and fabrication. However, the attempts to make biotechnology more similar to engineering

have lead synthetic biologists to distinguish design as an autonomous phase. This entails both a higher division of labor between design and wet lab experimental work and the development of methods and tools for design. Subsequently, design tools such as the use of computer-assisted design and standardization have become more prevalent. Following classical engineering, design is recognized as a distinct and creative form of problem-solving that does not aim to find one correct solution, but to assess between multiple possible solutions based on pragmatic criteria, such as efficiency, reliability, cost, or even the aesthetics of design (Kroes 2012).

The autonomous design phase produces its own conceptual questions concerning how to model and design new systems. Similar to engineering, synthetic biologists have proceeded to systematize their design methodology and develop distinct approaches to achieve reliable bioengineering of organisms with desirable traits. A systematic design approach should define the methods, heuristics and performance criteria that the engineer uses to solve design problems and to assess the success of the design. The most prominent design approach in synthetic biology is rational design, which closely resembles the traditional design paradigm of engineering (see Kroes 2012). Subsequently, rational design is often promoted and discussed in the more engineering-dominated conceptions of synthetic biology that emphasize analogies between engineered and living systems (Knuuttila and Loettgers 2013). However, rational design is frequently complemented and sometimes even replaced by more biologically-inspired evolutionary methods, the most visible of which is directed evolution. Directed evolution relies on a very different conceptualization of its design tasks, and, as we will see in Section 3, treats biological variation very

differently in comparison to rational design. I will next give a short account of both of these design approaches and their central heuristic differences.

Rational design aims to solve its design problems by approaching them from the perspective of ‘proper characterization of parts’. Like the design methods of traditional engineering sciences, it is a forward-engineering method and its main principles are (i) decoupling of design tasks, (ii) the use of standardized modular components and (iii) hierarchical organization (Endy 2005; Heinemann and Panke 2006). *Decoupling* means the separation of design goals into sub-tasks that can be tackled separately. In a decomposable system, each of these design sub-tasks could be solved semi-independently in relation to other tasks. This is done by constructing devices and modules with corresponding sub-functions by using *standardized modular components*. With rational design, characterization of a set of standardized components that can be predictably assembled to form more complex systems is one of the main goals of synthetic biology (Endy 2005). Possible standardized DNA components can be reporter genes, ribosome binding sites, promoter DNA sequences, and so on. Use of a standardized toolbox of parts takes its inspiration from electric engineering, and many of its standardized parts are characterized through similar language, as synthetic biologists talk, for example, about genetic switches and circuits (Purnick and Weiss 2009). Finally, designed systems should exhibit *hierarchical organization*, where standardized components form the foundational level of bottom-up system construction. Standardized components are put together to form more complex modules, such as bistable genetic switches or regulatory circuits (Andrianantoandro et al. 2006). In turn, these “second-level systems” should be able to be

combined to form more autonomous systems, such as whole networks (Purnick and Weiss 2009). In other words, *parts make devices and devices make systems*.

In these circumstances, the main task in rational design is to carefully specify, choose, and combine the individual components of the system resulting in a predictable composite behavior (Cambray et al. 2011; Haseltine and Arnold 2007). Rational design is often complemented or replaced with other design approaches, which are inspired by natural evolution. From the more biologically-inspired design methods, the most widely-used evolutionary design approach is *directed evolution*, which aims to use random mutations and deliberate selection to guide the designed system to evolve in a desired direction (Blake and Isaacs 2004). Design through directed evolution proceeds by iterating between two stages: the *library generation* and the *selection/screening* stage (Marguet et al. 2007). In the first stage, random mutations are induced into target sequence (e.g., through error-prone replication, PCR errors or DNA shuffling) to create a library of mutants that exhibit relevant variation. In the second stage, the generated library is screened and one or more mutants, with desired phenotypic differences, are selected for further iterations. Library-generation and a screening/selection cycle is repeated until a desired endpoint is achieved. Therefore, the work of a synthetic biologist using directed evolution resembles selective breeding in agriculture and animal husbandry (Lewens 2013).²

² Directed evolution is framed as a more biologically inspired design method than rational design. However, one should not consider directed evolution as exclusively biological design approach that is incompatible with engineering principles. Many engineering projects also rely on similar trial and error design methods (see Calcott et al. 2015), or evolutionary based design, such as use of genetic algorithms.

Directed evolution relies on different heuristics and the conceptualization of its design tasks than rational design. As with natural evolution, it modifies and tinkers with pre-existing structures rather than designs and constructs systems by manipulating their parts. Subsequently, directed evolution is not limited to any certain level of hierarchy, and can be applied to modify both individual parts as well as more complex modules or whole systems and organisms (Marguet et al. 2007). Therefore, it can be used to optimize or tune non-functional systems that do not exhibit the desired functionality (Yokobayashi et al. 2002). One such target for tuning by directed evolution are those rationally designed systems that do not achieve the intended functionality, or that change their functionality due to context-sensitivity. I will discuss the latter case and the effects of environment on the variation of functionality in Section 3.3.

This redesign of rationally-designed systems by directed evolution should exemplify the complementary role of different design approaches and their respective design principles (Dougherty and Arnold 2009). Both design approaches have their benefits and limitations. Whereas rational design relies on the decomposability of the system and the availability of well-characterized standardized parts, directed evolution is primarily limited by the ability of the researcher to induce mutational variation in the target sequence, and to detect and select desired phenotypes from the generated library of mutants. However, rational design is often framed as the primary design methodology of synthetic biology, and evolutionary design approaches such as directed evolution are considered to be only intermediate steps to overcome the current limitations of rational design (Bujara and Panke 2010). This follows the traditional engineering of idea conceptualizing design practice as a form of

rational problem-solving (see Simon 1996). Directed evolution, where the designer both induces random mutations into the system and cannot predict the exact direction in which the system will evolve, is, in turn, seen as a less rational form of design. Consequently, directed evolution is described both as an “irrational” (Lewens 2013, p. 642) and “semi-rational” (Guimaraes et al. 2013, p. 72) design approach, or is even characterized as “engineering without design” (Marguet et al. 2007, p. 610).

This difference between rational design and directed evolution can be made more explicit by considering it from the perspective of biological variation. The inherent stochasticity and messiness of biological systems imposes a major challenge for the attempts to engineer them, and the two design approaches provide different solutions to deal with variation. Based on this difference, Giese et al. (2013) categorizes synthetic biologists into two camps in relation to their stance to biological variation, namely *rational* and *evolutionary* groups. Rational group considers randomness to be a problematic phenomenon and biological variation as something to be suppressed. In turn, evolutionary group sees variation and randomness as “an elementary part of biological function and development,” and as something to be accommodated in the design solutions (Giese et al. 2013, p. 325).

To some extent, this juxtaposition between rational and evolutionary design mirrors the engineering and biology disciplinary roots of synthetic biology (see Cameron et al. 2014). Subsequently, the unpredictability and noisiness of evolving biological systems highlights one of the major disanalogies in relation to engineered systems (Knuuttila and Loettgers 2014). Several classical engineering fields have established standards and protocols that allow them to design and construct objects like bridges and mobile phones with high

predictability. Whereas biological systems are affected by evolutionary change and both internal and external noise, classical engineering fields have technical infrastructure and foundations that allow them to achieve predictable control of their systems. As Arkin and Fletcher (2006, p. 114.3) put it, “Thermal fluctuations that drive stochastic behavior can typically be ignored or managed in traditional engineering, but often not in cells.”

Conscious of this disanalogy, Agapakis and Silver (2009) note that in the engineering of biological circuits, designed synthetic systems are defined explicitly in opposition to evolved systems. Therefore, the question concerning the role of biological variation and randomness is not important simply to coax out the respective methodological differences between different design approaches of synthetic biology, but it relates to a larger question in the center of synthetic biology about the engineerability of biological systems. Indeed, to better understand these differences between “rational” and “evolutionary” groups of synthetic biologists, a more nuanced understanding of biological variation is required. I will next turn to my core question, which is the role and effects of biological variation, and how they relate to the requirements of design in synthetic biology.

3. Variation

I distinguish between three different types of variation. *Genetic variation* (Section 3.1.) refers to the variation in the nucleotide sequence, usually due to the accumulation of mutations. It is the most important type of variation for both design approaches, as genetic content is the primary target for manipulations, and most phenotypic change is caused by genetic change. In turn, I distinguish between two types of phenotypic variation.

Functional phenotypic variation (Section 3.2.) is defined as the variation of the intended

target phenotype under the same environmental conditions, whereas *environmental phenotypic variation* (Section 3.3.) refers to the variation of the target phenotype under different environmental conditions. This distinction allows me, for example, to discuss the context-sensitivity of biological systems separately from other challenges of phenotypic variation to design approaches, such as, does each iteration of design cycle of directed evolution produce selectable differences in phenotypes.

This threefold typology follows the one used by Tawfik (2010), who discusses the different aspects of biological “messiness.” He considers biological variation from the evolutionary perspective: he is interested in the origin of messiness, whether evolution promotes things like heterogeneity and stochasticity, and how it affects evolutionary innovation and robustness. Although Tawfik (2010) does not discuss variation from the perspective of design or synthetic biology, I show that these distinctions are, in fact, relevant to synthetic biology. As Michel Morange (2013, p. 318) suggests, one should avoid “the false distinction drawn between the work of synthetic biologists and the action of evolution.” By applying the evolutionary discussion concerning variation to design of synthetic systems, we can better understand one of the core methodological controversies of synthetic biology.

Different types of variation provide distinct benefits and challenges at various stages of engineering, and these problems differ depending on the design approach that is used. For example, rational design tries to suppress the effects of environmental phenotypic variation by characterizing the functional ranges of its parts. Directed evolution, in turn, mitigates these effects by tuning its systems to the selective pressures of the implementation

environment. I will make two further distinctions to clarify these effects. First, I separate the engineering process into two distinct phases: the *design* phase and the *implementation* phase. In the design phase, synthetic biologists try to plan and model the system that exhibits the intended functions. It includes, e.g., *in silico* and *in vitro* research, as well as computer-aided modeling. Design in synthetic biology is usually done in a highly-controlled environment, and it is also the phase where the previously discussed design approaches are applied. In turn, implementation is the post-design phase when the synthetic biological system is used and maintained in its intended target environments. After implementation, systems are still susceptible to the effects of biological variation, as they undergo evolution, accumulate mutations and face environmental change. It should be noted that in most cases the design and the implementation phases are not fully separate. Synthetic biology projects often iterate between the design, implementation and redesign of their systems, and activities like prototype construction transcend the design/implementation divide. However, as I will show in the following sections, biological variation has distinctly different effects in the design and implementation phase. Furthermore, I will analyze how the ability of the designed system to control and maintain its functionality under this variation depends partly on the choice of the design approach. Second, I distinguish between *synchronic* and *diachronic* variation. Synchronic variation refers to the existing variation in a population or sets of populations at any one given time. In turn, diachronic variation denotes variation and change in time, usually during multiple

generations.³ This distinction allows me to tease out some additional differences between the design approaches.

3.1. Genetic Variation

Genetic variation is both the most important and the best-understood form of biological variation, due to the importance of DNA and genes in evolutionary and molecular biology. It is defined as variation in the nuclear sequence, due to mutations, insertions, deletions, and gene copy number variations (Tawfik 2010). Genetic variation is sometimes referred to as mutational variation (e.g., Lehner 2008; Rollié et al. 2012), as the accumulation of mutations is the main source of both desired and undesired genetic variation. Subsequently, these mutations affect the design process in different ways, depending on the design approach that is applied. The roles and effects of genetic variation are summarized in Table 1. Furthermore, to understand the utilization of genetic variation in directed evolution, it is important to make a distinction between genetic variation at one point of time and the accumulation of mutations in time, that is, between synchronic and diachronic genetic variation, respectively.

³ A synchronic/diachronic distinction can also be discussed in relation to design goals. In his analysis of software engineering, Calcott (2014, p. 298) defines a *synchronic goal* as the attempt to “make the software do something useful now,” and a *diachronic goal* means to “make the software easy to modify” in the future. As Calcott notes, analysis of engineering tends to focus on synchronic goals, whereas diachronic aspects are overlooked.

[Insert Table 1 about here]

Genetic variation, especially in bacteria, is often highly correlated with phenotypic variation. As most synthetic biology projects have practical goals that aim to achieve improved or novel functionality, synthetic biologists are primarily interested on the phenotypic effects of their design. It is, therefore, difficult to analyze the consequences and utilization of genetic variation without invoking its possible phenotypic effects. However, there are reasons to discuss genetic variation separately from phenotypic variation. In natural systems it provides the basis for evolution, whereas in biotechnology genetic content is the primary target for manipulations. Both rational design and directed evolution target their interventions on the genetic content, either through the use of DNA-based parts or induction of genetic variation. Secondly, though genotype and phenotype are correlated, the phenotypic effects of genetic variation are not uniform, and organisms generate phenotypic variation from genetic change in various ways (see Gerhart & Kirschner 2007; Tawfik 2010). Finally, decoupling genetic and phenotypic variation might provide a way to increase the mutational robustness of the system, allowing the designed system to maintain functionality after mutations (Kitano 2004). In Section 3.2. I will further discuss how the correlation of genetic and phenotypic variation affects different design approaches, but first, I will analyze the role of genetic variation in synthetic biology.

For rational design, most problems of genetic variation arise after the design work, in the implementation phase. Successful characterization and standardization of parts should allow rational design to plan and design its systems with predictable outcomes at the design phase, where many environmental factors and promiscuous interactions can be

controlled. Most unpredictable genetic variation in this phase is due to methodological limitations, e.g., technical and practical problems in the synthesis and assembly of a nucleotide sequence. This type of variation can be overcome with the development of less error-prone syntheses and proof-reading methods. However, the effects of synthesis errors should not be underestimated as this type of variation imposes challenges especially for more complex synthetic biology projects such as whole-genome engineering. The probability of error-free synthesis diminishes as the size of the synthesized genome increases. For example, in the case of *Mycoplasma laboratorium* genome engineering, a single synthesis error in one DNA fragment caused the synthetic systems to be non-functional, and led to significant delays in the project (Gibson et al. 2010).

In turn, induction and utilization of genetic variation is at the heart of directed evolution. Synchronic variation is not just left unsuppressed, it is a necessary factor whose emergence is encouraged in the library-generation stage of design, for example, by using error-prone replication or DNA shuffling (Marguet et al. 2007). This should create synchronic genetic variation, that is, a set of varying DNA molecules that can then be subjected to selection or high-throughput screening. If successful, iterating the library-generation and selection/screening cycle will lead to controlled diachronic genetic variation, where different generations of mutant lineages vary in their target sequence. However, the induction of variation is not a completely random process and successful application of directed evolution requires both types of variation to be controlled. In optimal cases, directed evolution requires synchronic genetic variation only in sequences that are relevant for target functionality. Unfortunately, other parts of the genetic sequence are susceptible

to mutations as well, and, subsequently, can cause the accumulation of undetected mutations. This unintended genetic variation hinders the independent screening of intended properties (Haseltine and Arnold 2007).

After the intended design is achieved, both rational design and directed evolution still face similar problems with genetic variation in the implementation phase. Control of the environment and its selective pressures is more difficult outside the laboratory setting of the design phase. In time, the DNA sequence of the designed system accumulates mutations, which leads to deterioration of the original designed system. Synthetic systems have, therefore, a tendency to accumulate diachronic genetic variation after they are implemented. This diachronic variation can lead to synchronic variation if multiple co-existing mutant lineages evolve from the designed sequence, or to a case where one mutant lineage replaces the original system. The probability of this depends on the selective pressures of the implementation environment and on the phenotypic effects of the genetic variation. It also shows why the correlation of genetic variation and phenotypic variation must be considered. If genetic and phenotypic variation are highly correlated, the system is sensitive either to selective pressures, or to changes in its functionality due to genetic deterioration (Wagner 2005). I will next turn to the question of phenotypic variation and its effects on designed systems.

3.2. Functional Phenotypic Variation

I distinguish two types of phenotypic variation: *functional* and *environmental*. Following a similar distinction given by Tawfik (2010), functional phenotypic variation is defined as

the variation of the intended target phenotype *under the same environmental conditions*. This variation might be quantitative (e.g., differences in the metabolic efficiency of a certain mutant) or qualitative (e.g., novel functionality or a loss of function).

Phenotypic variation is usually, at least partially, caused by differences in genotype, and therefore correlated with genetic variation. This means that changes in genetic variation often induce changes in phenotypic variation. However, this influence can also work the other way round. Phenotypic variation may guide genetic variation by influencing the possible evolutionary routes and fixing mutations through selection (Tawfik 2010).

Phenotypes of bacteria are especially sensitive to genetic changes. Even silent mutations, where a genetic change in the DNA codon does not change the amino acid that it codes, can have relevant phenotypic effects for synthetic biologists. Organisms have been observed to exhibit codon preference; synonymous codon replacement can affect translation efficiency and protein functionality (Angov et al. 2008). For example, Hu et al. (2013) tested 342 variants of antibody coding gene with synonymous codon usage. The produced proteins retained identical original amino acid sequence, but exhibited significant functional differences in solubility and antigen-binding affinity. This was explained by differences in the availability of corresponding transfer-RNAs for synonymous codons. Rare codons were translated at different rates which, subsequently, influenced the protein folding. This result shows how genetic variation can guide protein synthesis and affect protein phenotype without amino acid substitutions. On the other hand, the same genotypes can also exhibit different phenotypes in the same environment. This type of variation is due to developmental changes or stochastic events such as phenotypic noise (Tawfik

2010). Together, these discoveries demonstrate why the influence between genetic and phenotypic variation is not a straightforward question even in bacteria.

The stances of rational design and directed evolution towards functional variation are similar to their stances toward genetic variation discussed in the previous section. The effects of functional phenotypic variation are presented in Table 2. Due to the correlation of genetic and phenotypic variation, there is an important symmetry in the respective effects of genetic and functional variation to design. Synthetic biologists are concerned about the actual relevant functional effects that genetic variation might have. However, this correlation of genetic and phenotypic variation has different consequences for rational design and directed evolution.

[Insert Table 2 about here]

As discussed in the previous section, synthetic biologists are usually able to control most unintended variation in the design phase. For rational design, there should be no unintended synchronic functional variation in the design phase as the approach aims to make phenotypic variation predictable by testing, modeling and reassembling the systems. Predictable design is facilitated by the modularity of used components and the suppression of phenotypic noise. The main task of the design phase is to construct systems whose phenotypes exhibit most closely the intended functionality. Subsequently, possible functional variation usually arises from methodological problems, either when DNA synthesis and assembly errors have phenotypically detectable consequences, or when phenotypic noise cannot be suppressed.

In turn, directed evolution aims to achieve its design goals by utilizing both synchronic and diachronic functional variation. Successful application of directed evolution requires that in each iteration of the library generation stage, induced random mutations have phenotypic effects that allow the synthetic biologist to screen and select the desired phenotypes for further iterations. Relevant synchronic functional variation in each iteration of the design cycle is, therefore, necessary for successful application of directed evolution. Furthermore, to reach a desired phenotypic outcome, and to achieve the intended design goals, directed evolution requires a certain type of diachronic functional variation. There must be an incremental pathway of beneficial mutations from a starting point to a desired phenotype (Dougherty and Arnold 2009).

This shows an important tension concerning how the correlation of genetic and phenotypic variation affects the application of directed evolution. For directed evolution, each genotypic change should ideally bring about detectable changes in the phenotype. This allows each iteration of the library generation stage to produce enough phenotypic variation that can be screened or selected for improved functionality. Therefore, in its design phase, directed evolution benefits from a closer correlation of genotypic and phenotypic variation. However, as Zakeri and Carr (2015, p. 57) note: “once a successful design is achieved, we typically prefer that the designed system remain static, not deviating from the original specifications.” After that point, further variation can only disrupt the intended design.

Subsequently, both rational design and directed evolution benefit from mutational robustness in the implementation phase, that is, if the designed system can retain its

functionality even though being subject to genetic degradation and selection. There is an interest in robustness-promoting structures that provide a “genetic buffer” by decoupling genetic variation from the phenotype (Kitano 2004). For example, heat-shock chaperone proteins (HSPs) stabilize the structure of several proteins by aiding protein folding despite genetic changes and environmental stress. HSPs increase the robustness of the system and mask the effects of genetic variation by helping proteins to fold correctly when they have accumulated detrimental mutations (Rutherford 2003). However, if the correlation between genetic and phenotypic variation is higher, genetic change might lead to changes in functionality. In time, genetic change can cause diachronic phenotypic variation which – as harmful mutations are more probable than beneficial ones – can lead to a loss of functions (Haseltine and Arnold 2007). Even in cases where mutations lead to a novel functionality, these promiscuous functions are usually in opposition to the intended design goals, and can potentially be harmful to the environment or human health.

Possible synchronic functional variation also has important effects on implementation. As mentioned in Section 3.1., accumulation of mutations can lead to genetic divergence and the emergence of multiple mutant lineages. If this genetic variation has phenotypic effects with different fitness consequences in the implementation environment, synchronic functional variation can lead to competition and natural selection, and even to cases where the mutant lineage replaces the original designed strain. The latter poses a particular problem for designed systems, where the designed properties do not necessarily provide fitness benefits for the organism. Synthetic biologists have tried to estimate the genetic reliability of standardized parts by attaching a GFP-reporter gene to the device (Sleight et

al. 2010). By measuring the expression level of the reporter gene, synthetic biologists have been able to estimate parts' evolutionary half-life: the number of generations it takes for less than 50% of the population to have the designed functionality. Studies indicate that, for example, in the common *Escherichia coli* host, typical genetic parts lose their reliability after ~100 generations (Renda et al. 2014). This is explained by the fact that synthetic biology devices often cause a fitness cost to their host organisms. Systems with designed genetic devices impose a metabolic burden, as they usually require cells to synthesize additional RNAs and proteins that are unnecessary for the host survival, or that interfere with native cellular processes (Sleight et al. 2010). Subsequently, deleterious mutations that inactivate the designed genetic parts may be favored by selection. This might result in the dominance of non-productive mutants, as the estimates of evolutionary half-life of genetic parts suggest.

Rational design, in contrast, generally benefits from a lower correlation of genetic and phenotypic variation. This increases the chance that post-implementation genetic change does not cause relevant phenotypic changes. In other words, it increases the likelihood of mutations being neutral, which enables the system to maintain its stability and functionality. In turn, directed evolution benefits from a higher correlation of genetic and phenotypic variation in the design phase, but it faces similar problems of system “detuning” and lowered functionality after implementation. Directed evolution seems to face a trade-off between design efficiency and post-implementation system robustness. However, Wagner (2008) addresses this seemingly antagonistic relationship between robustness and evolvability by turning the analysis of robustness from genotype to

phenotype. Even though robust genotypes exhibit low evolvability, robust phenotypes actually promote evolvability. Wagner explains this by the fact that several alternative genotypes can produce the same phenotype; populations with robust phenotypes are more diverse and can access greater amounts of variation. Therefore, the effects of mutational robustness to the design process of directed evolution are not unequivocal, and depend, for example, on the fitness landscape of the manipulated system. Subsequently, directed evolution has been considered especially suitable for protein engineering, as proteins exhibit phenotypic robustness due to previously discussed folding stabilizing factors (Romero and Arnold 2009).

Finally, both design approaches aim to suppress functional variation in implementation. Rationally designed systems aim at structural solutions that increase genetic robustness, such as gene duplication (“back-up systems”), modular design or alternative metabolic routes (see Wagner 2005). In turn, directed evolution can decrease the chance of detuning by lowering the effects of the implementation environment’s selective pressures. This can be done by choosing a design environment that resembles the implementation environment, and therefore designing a system to function stably in the intended environment. As many systems of synthetic biology are sensitive to their implementation environment, this “environmental tuning” is an important rationale for the use of directed evolution. I will discuss this design goal in the next section when I analyze the effects of environmental phenotypic variation.

3.3. Environmental Phenotypic Variation

The second type of phenotypic variation, *environmental variation*, is defined as variation in the target phenotype *under different environmental conditions*. It describes how likely the system is to respond to changes in the environment and can thus be called its “context-sensitivity” (Güttinger 2013, p. 200) or “context-dependence” (Haseltine and Arnold 2007, p. 4). Context-sensitivity offers a major obstacle for proper implementation in synthetic biology. The goal of standardization is still far away, and in many cases achieving functionality even in one environment can require some post-hoc tinkering from synthetic biologists (O’Malley 2011; Purnick and Weiss 2009). Constructs of synthetic biology usually exhibit more environmental variation than natural systems, as they are less complex than their natural counterparts and include fewer redundant parts (“back-up” components) or robust structures. The ability for synthetic devices to work in more than one environment is often an important feature, and it has practical incentives, as context-sensitivity limits the application of systems, making them less cost-effective. Furthermore, the ideal of standardization requires a certain level of generality in parts, as standardized components are required to exhibit predictable behavior under a range of different environments (Endy 2005).

[Insert Table 3 about here]

As the variability of an environment is controlled in the design phase, environmental variation causes most of its challenges in the implementation phase. These challenges and their role in rational design and directed evolution are summarized in Table 3. Both design

approaches aim to predict and control the environmental variation in the implementation of their systems, though by different methods. Rational design aims to overcome environmental variation by characterizing the environmental responses and functionality ranges of its components. For each component, there should be some range of domains and functional roles where the component could be predictably used. This ideal has been successful for a limited number of simple designs in rational design, but not for more complex structures. In many cases, rational design constructs exhibit unpredictable environmental variation that still cannot be mitigated by standardization, and which limits the generality of its constructs (Purnick and Weiss 2009).

For directed evolution, environmental variation is also a challenge. However, directed evolution has better ways of mitigating the effects of environmental variation. Directed evolution allows synthetic biologists to circumvent the effects of environmental variation by finding particular solutions for each environment. Multiple iterations of the library generation and screening/selection cycle allow synthetic biologists to carefully choose mutants that exhibit higher functionality in the given environment (Haseltine and Arnold 2007). This ability to tune the non-functional system to achieve functionality in a novel environment is one of the main rationales for the use of directed evolution.

However, there is a trade-off as the gained functionality is usually achieved with higher context-specificity and decreased generality, as new mutants are tuned to work only in their specified environment. Therefore, directed evolution leads to an increased coupling between the system and its environment. In this way, use of evolutionary design methods resembles the use of genetic algorithms that, similarly, off-load some of the problem

solving to the environment (see Kuorikoski and Pöyhönen 2013). Due to the increased coupling between the system and the environment, environmentally tuned systems are harder to implement into a new domain. Environmental variation causes problems to both design approaches in synthetic biology, but directed evolution can better mitigate those problems with the loss of generality. Therefore, directed evolution is often used to supplement rational methods to overcome environmental variation and find particular solutions, “kludges”, for each environment (see O’Malley 2011).

Sometimes this higher context-sensitivity can be beneficial, for example, in the case of trophic containment (Marliere 2009; Torres et al. 2016). In these experiments, bacteria are grown in a culture that include molecules (for example, synthetic coenzymes) that do not occur in their natural food chain, and directed evolution is used to tune bacteria to that environment. These “alien nutrients” form a selective medium and bacteria are forced to evolve so that their metabolism adapts to utilize the added compounds. The goal of the process is to develop a mutant lineage that requires the exogenous supply of synthetic nutrients (Marliere 2009). This should prevent bacteria from growing in a natural environment which lacks the necessary nutrients, thus decreasing the risk of accidental release of GM bacteria outside the laboratory. Subsequently, the increased context-sensitivity that is achieved by directed evolution is an actual design goal.

Finally, there is another important trade-off that affects both design approaches in a similar way. Phenotypic noise, caused, e.g., by unstable expression patterns of gene networks of stochastic developmental factors, can hinder successful and predictable design (Tawfik 2010). Engineered systems usually aim to suppress and minimize the effects of noise, as it

causes unpredictability and inefficiencies (Knuutila and Loettgers 2013). In most cases, the proper functioning of synthetic systems requires high fidelity and specificity of their interactions, such as the enzyme-substrate reactions of the metabolism. The fidelity of interactions is costly, as systems require structures that can reduce noise or discern a signal (Tawfik 2010). Therefore, dealing with noise causes design trade-offs between efficiency and metabolic cost, or between specificity of interactions and robustness of the system to perturbations, e.g., in structural conformations of enzymes. Additional noise-suppressing structures, such as proof-reading mechanisms and developmental constraints accrue additional metabolic costs. Furthermore, such mechanisms can lower the ability to induce random mutations in systems, which is a required design step in directed evolution.

However, even though phenotypic noise can cause unpredictability and impose fidelity costs, it has also been shown to play an important functional role in some cellular phenomena, such as regulation and differentiation (Eldar and Elowitz 2010). Suppression of noise can affect the functional ranges of the designed systems, and decrease their robustness or adaptability. Subsequently, even rational design can in certain circumstances benefit from variation, and allowing certain amount of noise can actually make systems more robust. It provides another example of the challenge that synthetic biologists face when trying to apply engineering principles and methodology to the design of biological systems.

4. Conclusions

This paper offers an analysis on the role of biological variation in synthetic biology and explains how different types of biological variation affect the design at given points in the engineering process. I have argued that a more fine-grained analysis of different types of variation allows us to better understand this juxtaposition and how variation actually is suppressed and utilized in the design of synthetic biological systems. In general, variation is more problematic for the rational design approach and, in turn, essential for the evolutionary approach. However, not even directed evolution embraces all instances of variation, for it must suppress certain types of variation or mitigate their effects. Rational design is prone to changes in time (genetic deterioration and detuning), and context-specificity. In turn, directed evolution requires closer correlation between genotype and phenotype, and loses generality in dealing with environmental variation. Some types of variation, such as phenotypic variation after implementation, propose challenges to both design approaches. Finally, all design approaches must consider of the effects of biological noise, as it accrues fidelity costs and causes subsequent design trade-offs.

Variation poses important challenge for design in synthetic biology, and its role cannot be reduced to the role that variation plays in evolutionary change. Neither should biological variation be approached only from the engineering perspective, and seen merely as a source of unpredictability and something to be suppressed. Rather, different types of variation and their effects offer an opportunity to understand the methodological foundations of synthetic biology, and its attempts to reconcile its biological and engineering roots.

Acknowledgements

This research has been supported by the Academy of Finland research project “Biological Knowledge through Modeling and Engineering: Epistemological and Social Aspects of Synthetic Biology” (PI: Prof. Tarja Knuuttila, Grant number 272604), Finnish Cultural Foundation and Finnish Centre of Excellence in the Philosophy of the Social Sciences. I am grateful to Alkistis Elliott-Graves, Rami Koskinen, Tarja Knuuttila, Jaakko Kuorikoski, Uskali Mäki, Jani Raerinne, Anita Välikangas, and referees for this journal that provided helpful comments on previous drafts of this paper

References

- Agapakis CM, Silver PA (2009) Synthetic biology: Exploring and exploiting genetic modularity through the design of novel biological networks. *Mol Biosyst* 5:704–713. doi: 10.1039/B901484E
- Andrianantoandro E, Basu S, Karig DK, Weiss R (2006) Synthetic biology: New engineering rules for an emerging discipline. *Mol Syst Biol* 2:1–14. doi: 10.1038/msb4100073
- Angov E, Hillier CJ, Kincaid RL, Lyon JA (2008) Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLOS ONE* 3:e2189. doi: 10.1371/journal.pone.0002189
- Arkin AP, Fletcher DA (2006) Fast, cheap and somewhat in control. *Genome Biol* 7:114. doi: 10.1186/gb-2006-7-8-114
- Blake WJ, Isaacs FJ (2004) Synthetic biology evolves. *Trends Biotechnol* 22:321–324. doi: 10.1016/j.tibtech.2004.04.008
- Bujara M, Panke S (2010) Engineering in complex systems. *Curr Opin Biotech* 21:586–591. doi: 10.1016/j.copbio.2010.07.007
- Calcott B (2014) Engineering and evolvability. *Biol Philos* 29:293–313. doi: 10.1007/s10539-014-9425-3
- Calcott B, Levy A, Siegal ML, et al (2015) Engineering and biology: Counsel for a continued relationship. *Biol Theory* 10:50–59. doi: 10.1007/s13752-014-0198-3
- Cambray G, Mutalik VK, Arkin AP (2011) Toward rational design of bacterial genomes. *Curr Opin Microbiol* 14:624–630. doi: 10.1016/j.mib.2011.08.001
- Cameron DE, Bashor CJ, Collins JJ (2014) A brief history of synthetic biology. *Nat Rev Microbiol* 12:381–390. doi: 10.1038/nrmicro3239
- Dougherty MJ, Arnold FH (2009) Directed evolution: New parts and optimized function. *Curr Opin Biotech* 20:486–491. doi: 10.1016/j.copbio.2009.08.005
- Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. *Nature* 467:167–173. doi: 10.1038/nature09326
- Endy D (2005) Foundations for engineering biology. *Nature* 438:449–453. doi: 10.1038/nature04342
- Gerhart J, Kirschner M (2007) The theory of facilitated variation. *Proc Natl Acad Sci U S A* 104:8582–8589. doi: 10.1073/pnas.0701035104

- Gibson DG, Glass JI, Lartigue C, et al (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329:52–56. doi: 10.1126/science.1190719
- Giese B, Koenigstein S, Wigger H, et al (2013) Rational engineering principles in synthetic biology: A framework for quantitative analysis and an initial assessment. *Biol Theory* 8:324–333. doi: 10.1007/s13752-013-0130-2
- Guimaraes JC, Liu CC, Arkin AP (2013) From biological parts to circuit design. In: Zhao H (ed) *Synthetic biology: Tools and applications*. Elsevier, Amsterdam, pp 63–78
- Güttinger S (2013) Creating parts that allow for rational design: Synthetic biology and the problem of context-sensitivity. *Stud Hist Philos Biol Biomed Sci* 44:199–207. doi: 10.1016/j.shpsc.2013.03.015
- Haseltine EL, Arnold FH (2007) Synthetic gene circuits: Design with directed evolution. *Annu Rev Biophys Biom* 36:1–19. doi: 10.1146/annurev.biophys.36.040306.132600
- Heinemann M, Panke S (2006) Synthetic biology: Putting engineering into biology. *Bioinformatics* 22:2790–2799. doi: 10.1093/bioinformatics/btl469
- Heinemann M, Panke S (2009) Synthetic biology: Putting engineering into bioengineering. In: Fu P, Panke S (eds) *Systems biology and synthetic biology*. Wiley, New York, pp 387–409
- Houkes W, Vermaas PE (2010) *Technical functions: On the use and design of artefacts*. Springer, Dordrecht
- Hu S, Wang M, Cai G, He M (2013) Genetic code-guided protein synthesis and folding in *Escherichia coli*. *J Biol Chem* 288:30855–30861. doi: 10.1074/jbc.M113.467977
- Kitano H (2004) Biological robustness. *Nat Rev Genet* 5:826–837. doi: 10.1038/nrg1471
- Knuuttila T, Loettgers A (2013) Basic science through engineering? Synthetic modeling and the idea of biology-inspired engineering. *Stud Hist Philos Biol Biomed Sci* 44:158–169. doi: 10.1016/j.shpsc.2013.03.011
- Knuuttila T, Loettgers A (2014) Varieties of noise: Analogical reasoning in synthetic biology. *Stud Hist Philos Sci* 76–88. doi: 10.1016/j.shpsa.2014.05.006
- Kroes P (2012) *Technical artefacts: Creations of mind and matter: A philosophy of engineering design*. Springer, Heidelberg
- Krohs U, Bedau MA (2013) Interdisciplinary interconnections in synthetic biology. *Biol Theory* 8:313–317. doi: 10.1007/s13752-013-0141-z
- Kuorikoski J, Pöyhönen S (2013) Understanding nonmodular functionality: Lessons from genetic algorithms. *Philos Sci* 80:637–649. doi: 10.1086/673866

- Lehner B (2008) Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol Syst Biol* 4:170. doi: 10.1038/msb.2008.11
- Lewens T (2013) From bricolage to BioBricks™: Synthetic biology and rational design. *Stud Hist Philos Biol Biomed Sci* 44:641–648. doi: 10.1016/j.shpsc.2013.05.011
- Marguet P, Balagadde F, Tan C, You L (2007) Biology by design: Reduction and synthesis of cellular components and behaviour. *J R Soc Interface* 4:607–623. doi: 10.1098/rsif.2006.0206
- Marliere P (2009) The farther, the safer: A manifesto for securely navigating synthetic species away from the old living world. *Syst Synth Biol* 3:77. doi: 10.1007/s11693-009-9040-9
- Morange M (2013) Comparison between the work of synthetic biologists and the action of evolution: Engineering versus tinkering. *Biol Theory* 8:318–323. doi: 10.1007/s13752-013-0134-y
- O'Malley M (2011) Exploration, iterativity and kludging in synthetic biology. *C R Chim* 14:406–412. doi: 10.1016/j.crci.2010.06.021
- Packer MS, Liu DR (2015) Methods for the directed evolution of proteins. *Nat Rev Genet* 16:379–394. doi: 10.1038/nrg3927
- Purnick PEM, Weiss R (2009) The second wave of synthetic biology: From modules to systems. *Nat Rev Mol Cell Biol* 10:410–422. doi: 10.1038/nrm2698
- Renda BA, Hammerling MJ, Barrick JE (2014) Engineering reduced evolutionary potential for synthetic biology. *Mol Biosyst* 10:1668–1678. doi: 10.1039/C3MB70606K
- Rollié S, Mangold M, Sundmacher K (2012) Designing biological systems: Systems engineering meets synthetic biology. *Chem Eng Sci* 69:1–29. doi: 10.1016/j.ces.2011.10.068
- Romero PA, Arnold FH (2009) Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* 10:866–876. doi: 10.1038/nrm2805
- Rutherford SL (2003) Between genotype and phenotype: Protein chaperones and evolvability. *Nat Rev Genet* 4:263–74. doi: <http://dx.doi.org/10.1038/nrg1041>
- Silver PA, Way JC, Arnold FH, Meyerowitz JT (2014) Synthetic biology: Engineering explored. *Nature* 509:166–167. doi: 10.1038/509166a
- Simon HA (1996) *The Sciences of the Artificial*, 3rd ed. MIT Press, Cambridge, MA
- Sleight SC, Bartley BA, Lieviant JA, Sauro HM (2010) Designing and engineering evolutionary robust genetic circuits. *J Biol Eng* 4:12. doi: 10.1186/1754-1611-4-12

- Tawfik DS (2010) Messy biology and the origins of evolutionary innovations. *Nat Chem Biol* 6:692. doi: 10.1038/nchembio.441
- Torres L, Krüger A, Csibra E, et al (2016) Synthetic biology approaches to biological containment: Pre-emptively tackling potential risks. *Essays Biochem* 60:393–410. doi: 10.1042/EBC20160013
- Wagner A (2005) *Robustness and evolvability in living systems*. Princeton University Press, Princeton, NJ
- Wagner A (2008) Robustness and evolvability: a paradox resolved. *Proc R Soc B* 275:91–100 . doi: 10.1098/rspb.2007.1137
- Yokobayashi Y, Weiss R, Arnold FH (2002) Directed evolution of a genetic circuit. *Proc Natl Acad Sci U S A* 99:16587–16591. doi: 10.1073/pnas.252535999
- Zakeri B, Carr PA (2015) The limits of synthetic biology. *Trends Biotechnol* 33:57–58. doi: <http://dx.doi.org/10.1016/j.tibtech.2014.10.008>

GENETIC VARIATION		
Variation in the genetic content (nucleotide sequence) due to, e.g., mutations, insertions, deletions and gene copy number variation.		
Design Phase	<u>Rational design</u>	Synthesis and assembly errors.
	<u>Directed evolution</u>	Indispensable part of design. <i>Synchronic</i> : Generation of mutant library. <i>Diachronic</i> : Accumulation of incremental mutations.
Implementation Phase		<i>Diachronic</i> : Accumulation of unintended mutations (Genetic degradation). <i>Synchronic</i> : Emergence of multiple mutant lineages.
Suppression of variation	<u>Rational design</u>	Development of error-free synthesis; RNA/DNA repair and proof-read mechanisms.
	<u>Directed evolution</u>	Development of methods that induce mutations only in the target sequence.

Table 1. Role and effects of genetic variation on design.

FUNCTIONAL PHENOTYPIC VARIATION		
Variation in the phenotypes of studied systems <i>under the <u>same</u> environmental conditions.</i>		
Design Phase	<u>Rational design</u>	Synthesis and assembly errors with phenotypic effects.
	<u>Directed evolution</u>	Indispensable part of design, required for selecting improved functionality. <i>Synchronic</i> : Selectable differences in phenotypes <i>Diachronic</i> : Incremental pathway of improved phenotypes.
Implementation Phase		<i>Diachronic</i> : Systems evolve to match the selective pressures of the environment. <i>Synchronic</i> : Emergence of phenotypically different mutant lineages (with possible fitness differences)
Suppression of variation	<u>Rational design</u>	Robust system design: e.g., modularity, “back-up” systems.
	<u>Directed evolution</u>	Environmental tuning.

Table 2. Role and effects of functional phenotypic variation on design.

ENVIRONMENTAL PHENOTYPIC VARIATION		
Variation in the phenotypes of studied systems <i>under <u>different</u> environmental conditions.</i>		
Design Phase		<i>[None, environmental conditions are controlled in the design phase]</i>
Implementation Phase		<i>Synchronic: Unpredictable functionality in different environments (context-sensitivity). Diachronic: Loss of functionality or emergence of promiscuous functions in time.</i>
Suppression of variation	<u>Rational design</u>	Standardization and characterization of functional ranges in different environments.
	<u>Directed evolution</u>	Environmental tuning.

Table 3. Role and effects of environmental phenotypic variation on design.