VIDIUM I WHITE PAPER I No. 2 I December 2021

VIDIUM
SYSTEMS BIOLOGY SOLUTIONS

ARTICLE BY
Arnaud Bonnaffoux I Sara Munoz I Elsa Guillot

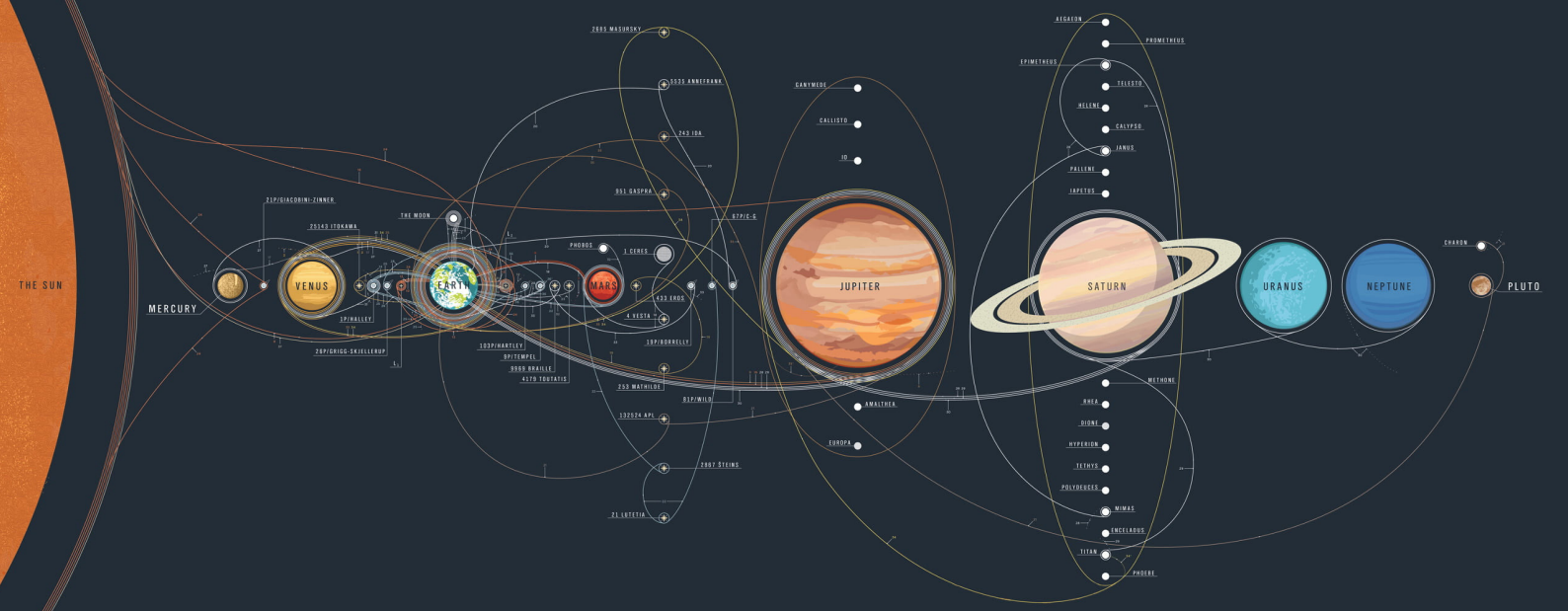
BIOLOGY FROM AN ENGINEERING POINT OF VIEW

HOW WOULD AN ENGINEER "FIX" A CELL?

The article entitled "Can a biologist fix a radio?" from Lazebnik in Cancer Cell, exemplifies the difficulty to analyze and fix a system from the descriptive approach in biology. Almost 20 years after the publication of this paper, new experimental data at the single-cell level reveal a new level of variability and heterogeneity which makes the picture even more complex. This second white paper is the engineer's answer to Lazebnik's paper: How would an engineer fix a cell? More specifically we will focus on the inference of Gene Regulatory Networks.

Image by Evan Ingersoll & Gael McGill "Cellular Landscape Cross-Section Through A Eukaryotic Cell" In https://t.co/mVs9dmh5bH

THE GRAIL : FINDING BIOLOGICAL RULES



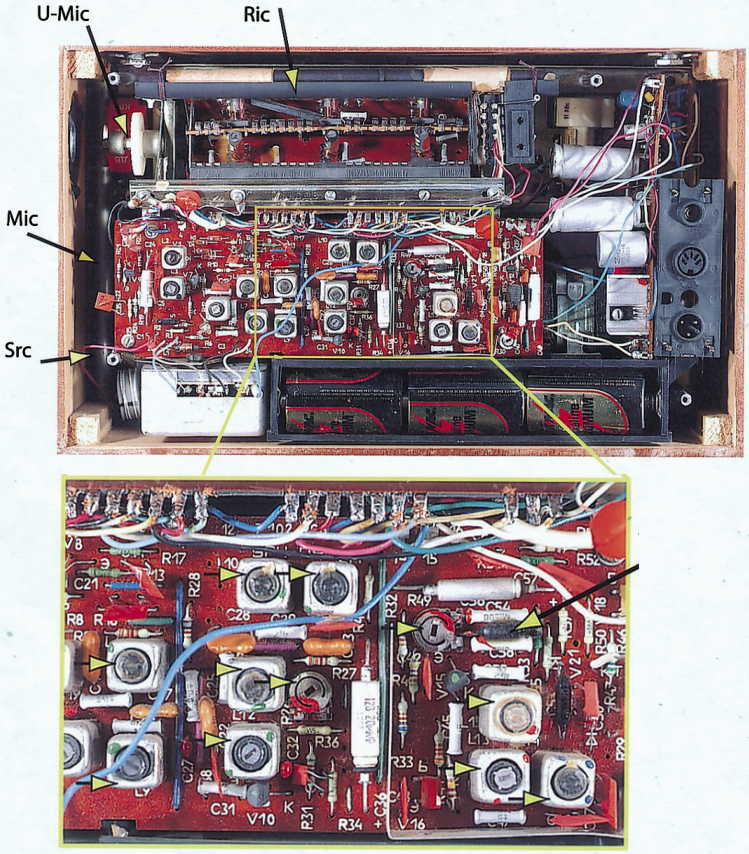
As stated in the previous Vidium White Paper entitled “Challenges, limits and opportunities in systems biology“, biologists are facing the big challenge of complexity. The recent blooming of single-cell assay technologies provides an unprecedented opportunity. However, this data is exploited by means of descriptive approaches: clustering, gene ontology, functional annotation, correlation, etc. This process is necessary but not sufficient to reach the goal of any experimental science: to find the underlying fundamental rules that drive system complexity. In astronomy, another experimental science, Newton was able to propose a simple model, the law of gravity, to explain the complexity of planet trajectories observed in the sky. This example illustrates how a simple generic law (gravity is constant) becomes an underlying fundamental rule generating complex and dynamic planet movements. Hence, Newton managed to explain the system complexity of planet trajectories.

While the rule is fixed for all planetary systems - we will use the term "invariant" to describe the rules - the movements are changing depending on each planet - we will use the term "variants" to describe the trajectory observations.

Between the 19th and 20th centuries, Physics successfully defined an incredibly limited set of four fundamental rules that are responsible for all physical phenomena observed in the Universe, including gravity [1].

Following the discoveries by English physicist Francis Crick in the 1950s, Biology defined the “central dogma” of DNA, a potential invariant. Once the Human genome was sequenced, biologists thought momentarily that they would be able to fully explain how human cells work. Indeed, DNA is a perfect candidate for the invariant rule of the human system, while RNA and protein are just variant products. However, as we all know, the genetic code by itself is far from sufficient in explaining the gene expression regulatory mechanisms that drive cell phenotypes and the complex mechanisms underlying the behavior of a complex living system.

THE LANGUAGE OF BIOLOGIST AND ENGINEERS



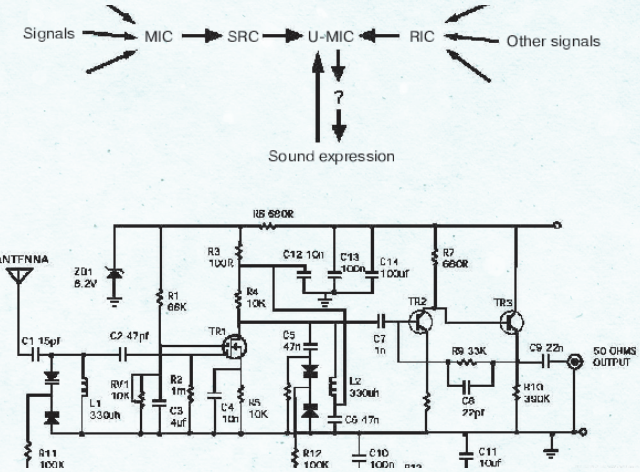


Fig.1. The internal components of a radio interpreted by a biologist (left) and an engineer (right). The horizontal arrows indicate tunable components of the radio (left). Image extracted from [2] and [3].

Biologists usually focus on describing, classifying, and correlating observations. They use descriptive diagrams to make qualitative interpretations that lack a quantitative approach. This limits the predictive or investigative value of the results to a very limited domain of applications.

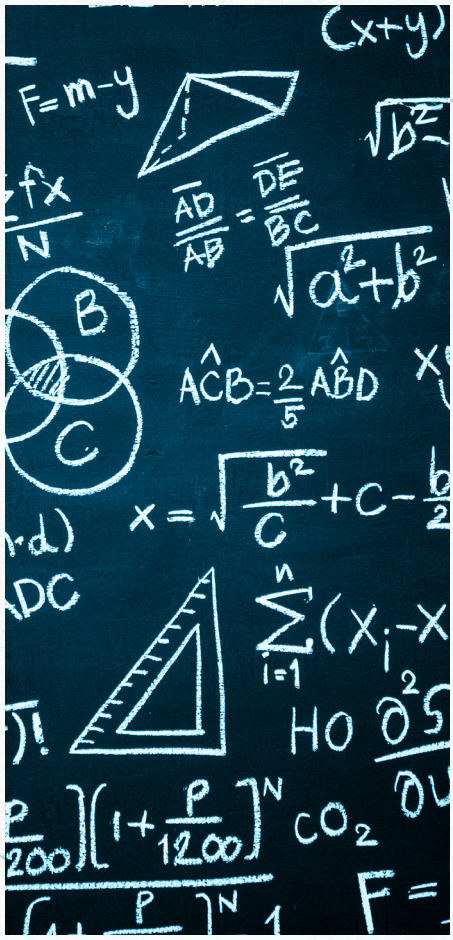
On the contrary, engineers communicate through standards: all the elements and connections are described according to the invariants of the system. Formalization is their key to communication. All trained engineers can unambiguously understand a quantitative diagram, where all components are regulated by key parameters. Another major benefit of formalization is that, from the description of the rules of the system, the model can be easily translated into an executable program that simulates the evolution of variants (often called states).
Returning to the example of the radio given by Lazebnik [2], biologists will open the radio then observe and describe a complete set of objects of various shapes, colors, and sizes, with some of them being required to function, as well as a set of connections between these objects (Fig. 1). However, even when the radio is reconstructed with all its components and all the corresponding connections wired, the radio may not work. This is because some of the components are tunable, for example a variable condenser. It might be that some are not tuned properly. Hence, in the absence of quantitative key parameters to evaluate these variables, the radio will remain non-functional.

To get out of this situation, where the activities of biologists are often limited to describing and correlating what they observe, **it is necessary to find the "invariants" of the system,** as is the case in other experimental sciences.

FORMALIZATION WITH STATES AND STRUCTURES

STATES & STRUCTURE, TWO FUNDAMENTAL CONCEPTS TO FORMALIZE BIOLOGICAL SYSTEMS

TO ADDRESS THIS CHALLENGE OF FINDING RULES UNDERLYING VARIABLE OBSERVATIONS, IT IS IMPERATIVE TO CORRECTLY FORMALIZE VARIANTS AND INVARIANTS IN BIOLOGY. TO DO SO, WE CAN USE THE ENGINEERING CONCEPTS OF STATE (VARIANTS) AND STRUCTURE (INVARIANTS) (WHICH COME FROM CONTROL THEORY).



Let's introduce some basic engineering notions of the **"states"** and **"structure"** of a dynamic system. The "states" are variables to describe the system at a given time. They can be measured like the coordinates of the planets in the sky.

In biology, the notions of phenotype and genotype mirror, respectively, the notions of "states" and "structure" used in engineering.

States are time-dependent variables that describe the system at a given time. The notion of structure refers to internal rules, laws, or interactions that drive the evolution of states. **The structure is time-invariant.** States and structures are usually formalized in a set of ordinary differential equations (ODE) to describe the dynamic evolution of a system over time.

In Control Theory [4], a system is modeled by a function F (the structure) which integrates **external stimuli (vector U)** to modify its **internal states (vector X).**

The function **F defines the system dynamic.** The dynamic of the system is described by the following ODE:

$$\frac{dX(t)}{dt} = F(X(t), U(t))$$

The use of a **differential equation as a function of time** shows the importance of the historical information received by the system to define its state at a given time T. Indeed, to know the state of a system at a time T, it is necessary to proceed to a temporal integration of the stimuli and its state:

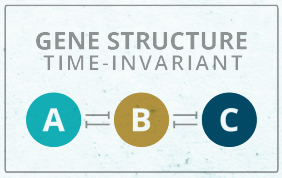
$$X(T) = \int_0^T F(X(t), U(t))dt$$

This temporal integration also means that **the state of the system behaves as dynamic memory.** When no more stimulus act on the system, the state does not go back to zero, it memorizes its previous value.

We also notice that the dynamic of the system (defined by function F) is not limited to the stimulus. It also considers the internal state of the system (X). The main consequence is the invalidation of the prevalent deterministic point of view in biology, where genotype and environment solely define the phenotype. Instead, it is also necessary to consider the current phenotype to determine the next phenotype. Figure 2 exhibits a simple example of a cell with a simple gene regulatory network that can react differently to the same stimulus.

A GENE REGULATORY NETWORK IS A DYNAMIC "STATE MACHINE"

"THE STRUCTURE OF THE GENE REGULATORY NETWORK (GRN) IS THE SET OF POSSIBLE INTERACTIONS BETWEEN GENES. THIS IS ESSENTIALLY A DYNAMIC SYSTEM."



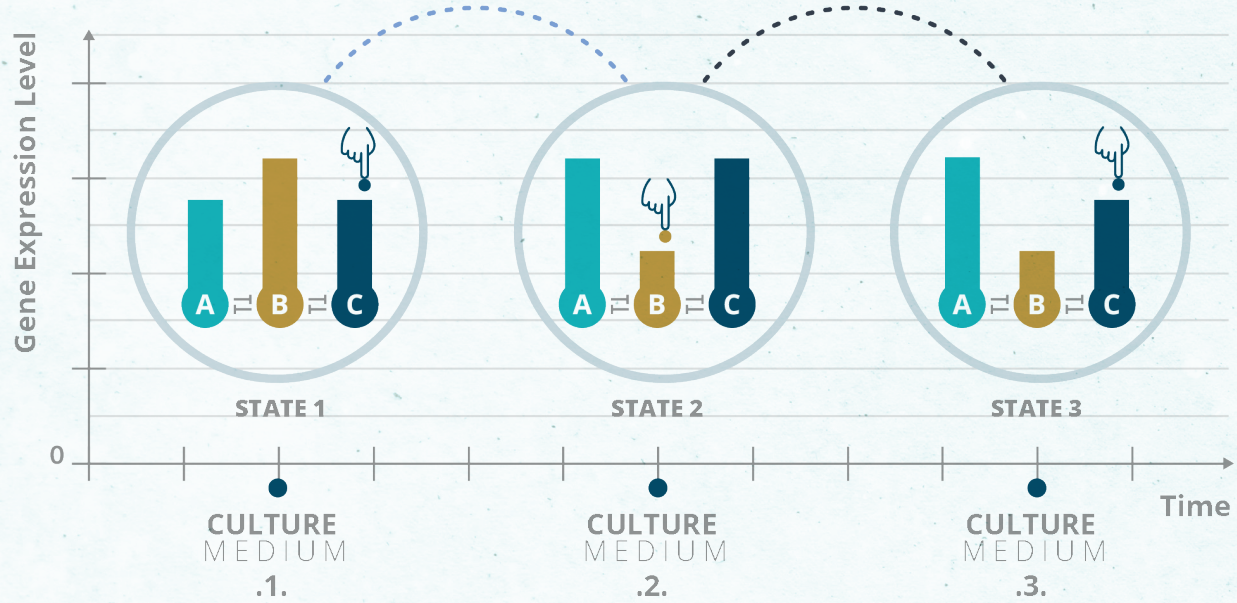


Fig. 2. The state evolution of a simple GRN depends on the history of its stimulus and states. Consider a simple GRN composed of 3 genes (A, B, & C) with mutual repressions. Suppose an experiment where initially the GRN is submitted to an external stimulus which represses gene C. Gene A is initially down and B is high. The expression level of each gene corresponds to the GRN state, and it is represented by a bar. In a first time, we remove the stimulus on C and add a stimulus on B, which represses gene B. Since A and C are no more repressed by gene B they are upregulated. In a second time, we remove the stimulus B and re-apply the stimulus on C to repress gene C. Gene A is still high because it has memorized its previous state and thus it still represses B. As we can see, the final state of the GRN (A+/B-/C-) is different from the initial state (A-/B+/C-), even if we apply the same stimulus to the same GRN.

For a cell, the expression level of any gene (RNA and protein amounts) can be considered as a state. Its expression and quantity vary through time, like during a differentiation process, and can be measured. The phenotype can be seen as a state.

Gene expression is regulated by complex interactions between molecules, that are directly or indirectly gene products. For example, transcription factors bind to DNA to promote or inhibit the expression of their target genes. All these interactions create a so-called Gene Regulatory Network (GRN), which we consider as the invariant structure.

A GRN corresponds to the set of possible interactions between genes as the one presented in Figure 2. The structure of a GRN includes the necessary conditions for each interaction to be effective. A GRN is constant over time and does not depend on the cell type, unlike the phenotype of a cell.

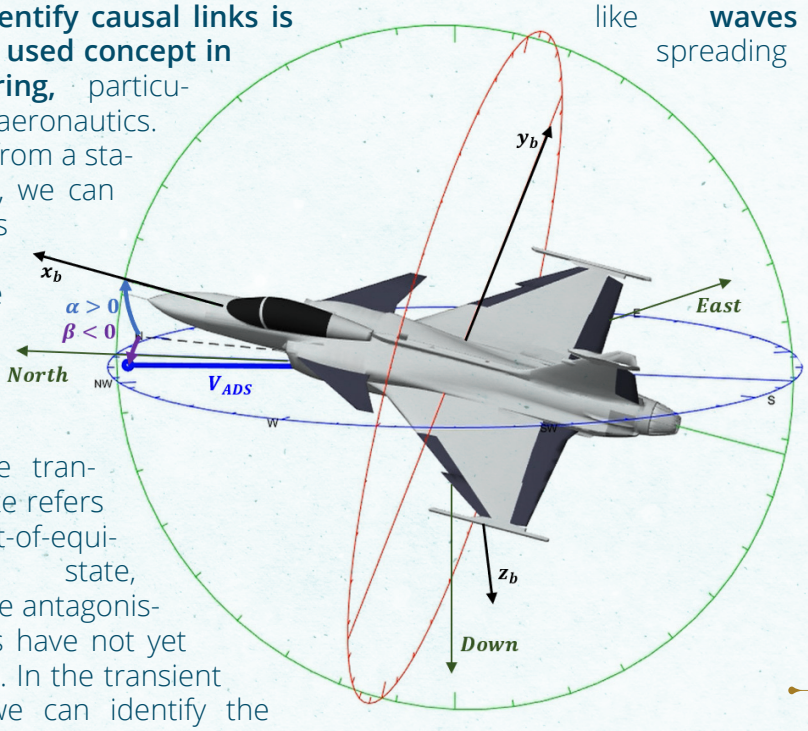
The example in Figure 2 emphasizes the importance to consider the transient regime, i.e. the history of stimulus and gene expression states, when trying to interpret the relation between phenotype (state), genotype (GRN) and environment (stimuli).

Thus, to consider the transient regime is essential when trying to infer system structure from states observation.

RETRO-ENGINEERING : FROM STATES OBSERVATION TO STRUCTURE INFERENCE

In the introduction, we stated that the Grail in biology is to find the biological invariants, like GRN structure. As we explained in the previous part, it is important to consider the **transient** regime to infer structure from the states. This is exactly how we proceed in engineering to retro-engineer systems, e.g., to find the rules of the system.

The study of transient dynamics to identify causal links is a widely used concept in engineering, particularly in aeronautics. Starting from a stable state, we can study its transient response by disturbing it with a stimulus. The transient state refers to an out-of-equilibrium state, where the antagonistic forces have not yet balanced. In the transient phase, we can identify the common dynamic variations (like oscillations) to a set of states, with their frequencies; we can also identify phase shifts that would betray causal links. In practice, **engineers use computational mechanistic models that implement the known rules** (aerodynamic forces for example). They estimate and tune these models' parameters (for example the aerodynamic coefficients Cx, Cz, ...) to fit and reproduce states oscillations that are observed during flight tests (Fig. 3).



During his PhD, A. Bonnaffoux applied this retro-engineering method to infer the interactions in the GRN. The first assessment is **to observe experimentally the transient behavior of gene regulation.** The idea is that, after a stimulation, genes regulate each others sequentially by cascade, like **waves** spreading through a network.

Fortunately, the typical time response for a gene (in superior eukaryotes) is several hours, which is easily accessible experimentally. Richard et al. [6] (see Fig. 7 from the article) observe these waves during chicken erythropoiesis. Based on a **mechanistic mathematical model of GRN** [7] A. Bonnaffoux has then developed the algorithm WASABI [8] **to infer the interaction parameters of the GRN.**

The most important advantage of inferring a GRN, based on the transient regime, is that we can split the inference problem and get rid of the curse of dimensionality [9] by reconstructing the GRN, adding genes one by one according to their wave time. This is an important contribution from the dynamic analysis approach towards solving the GRN inference problem.

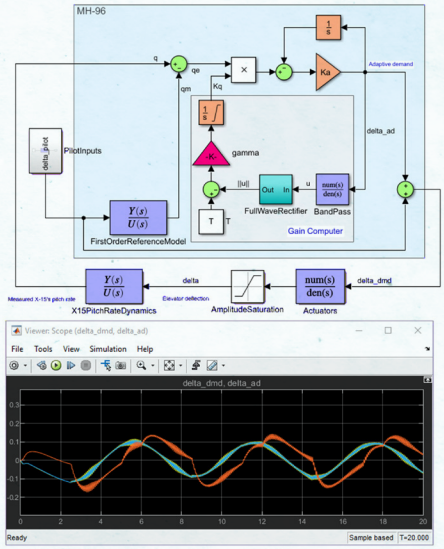


Fig. 3. The Simulink simulation environment of the X-15's pitch dynamics.
Image by the author Rodney Rodriguez [5].

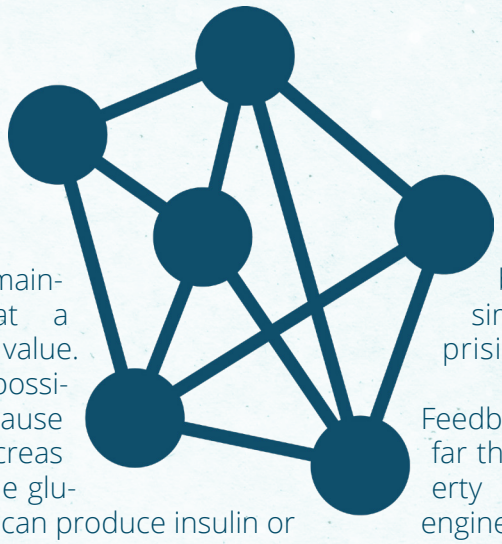
Since GRNs inferred from WASABI are directly interpretable (contrary to AI black boxes), we can link GRN topology to system behaviors (cell decision-making processes) and have a better understanding of **how complex behavior arise from typical patterns.**

FROM STRUCTURE TO BEHAVIOR : COMMON PATTERNS OF STRUCTURE IN ENGINEERING AND BIOLOGY

“ DURING MY CONVERSION IN BIOLOGY, I WAS SURPRISED TO RECOGNIZE MANY PATTERNS IN THE STRUCTURE OF CELLULAR PROCESSES THAT ARE SIMILAR TO THE ONES USED IN AN AIRCRAFT AUTOPILOT. AUTOPILOT IS AN AUTONOMOUS AND CRITICAL SYSTEM, WHICH ARE TWO IMPORTANT CHARACTERISTICS SHARED WITH BIOLOGICAL SYSTEMS THAT, FOR ME, ARE TIDILY LINKED TO THEIR ARCHITECTURE. ”

A. BONNAFFOUX

AUTONOMY & ROBUSTNESS



How does an autopilot work? An autopilot shall pilot an aircraft alone, it has the commands! The pilot just gives the objectives, i.e., the final airport destination, and (almost) the rest is managed by the autopilot itself. Basically, the core processes of an autopilot rely on **feedback loops.**

Let’s consider the example where the autopilot holds a required altitude (Fig. 4). To do so, we need sensors to estimate the current altitude, which is the state of our system. The estimated altitude is compared with the target altitude to generate a command to the actuators, which control the elevator governs. In response, the aircraft will change its altitude, which in turn will be detected by the sensor, closing the loop (Fig. 4).

In a biological system, this is equivalent to homeostasis. A certain state of a homeostatic system, such as glucose concentration, must be maintained at a specific value. This is possible because the pancreas senses the glucose and can produce insulin or glucagon to respectively decrease or increase glucose levels. The pancreas is the “autopilot” analog since it can sense, regulate and actuate on the glucose concentration. Then, both living and engineered systems use feedback loops to self-regulate.

In the middle of the last century, **Wiener et al. founded cybernetics [10], an aggregate of several fields like signal processing or control theory.** Nowadays, thanks to cybernetics principles, our engineered systems are way more complex, and we know far more about biological systems. Indeed, for someone, such as A. Bonnaffoux, who is familiar with both domains, the similarities are very surprising.

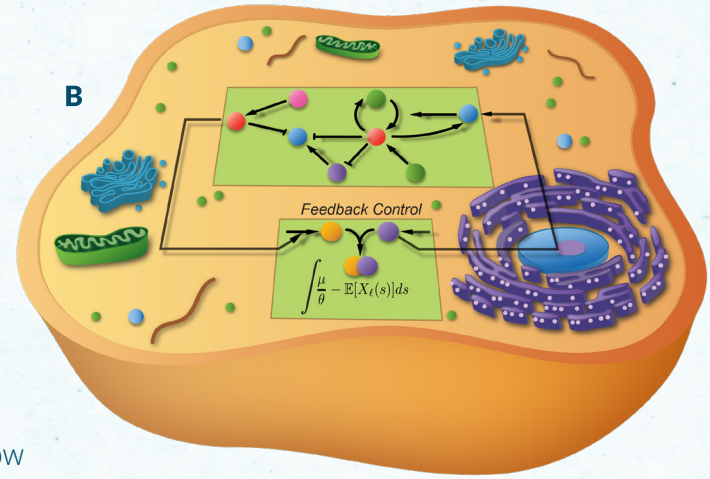
Feedback loops are not by far the only common property between living and engineering systems. For example, autopilots use extensively different kinds of filters (low pass, high pass, band stop, etc...) to process the sensor's recovered signal. For instance, high pass filters are used in engineering to remove biases or detect changes. In biology, the cell signaling pathways function similarly: they detect signal fold changes by using incoherent feedforward loops [11, 12]. Low pass filters are used in engineering to remove noise, which is what a cell does to distinguish signal from noise during its differentiation process [13]. Today, these patterns are used to engineer cells in the context of immunotherapy, to ensure robust cell behaviors [14].

CRITICALITY AND SAFETY

Cells and autopilot share another important property: they are critical. It means that a system failure could be a matter of life or death. When they are submitted to failures they are similar in their strategies to face this problem using the same structure.

i) The **apoptotic regulation** in cells is a well-conserved auto-destruction mecha-nism in multi-cellular organisms. Its main function is to ensure that if a cell is doing something wrong, it shall kill itself for the good of the organism, otherwise, it can become a malignant tumor cell. Autopilots have also an auto-disconnection system. If the autopilot became "crazy" and unable to disengage, it could crash the aircraft. That is why almost 50% of the autopilot code is dedicated to this function. If we look into details, there are further similarities, like the quality check for the detection of internal failures. This is performed by many "quality check points" in cells, e.g. checking the steps during the cell cycle or checking that a protein is well folded. An auto-pilot is also continuously checking its behavior.

ii) To ensure safety the key is **redundancy**: There is not a single autopilot computer but at least two to six computers monitoring each other. They all execute the same code, but if one computer disagrees, there is a vote between them and possibly a disconnection. The autopilot can also be monitored by other systems in the aircraft. Redundancy is a common architecture pattern observed in genetic. We know that critical genes are not alone and work in concordance with redundant genes. This is logical: the more critical a gene is, the more robust the system should be regarding its dependence to avoid a critical failure, and the simplest solution is to have redundant genes. This is exactly how we manage critical functions in aircraft. **The more critical a system is, the more redundancy is included.** If we unplug a critical system in an aircraft, we will see no effect on its behavior. This may be paradoxical, but it makes sense. How should we regard all the K.O. genetic screening performed in biology then? The current consensus interpretation is that if a K.O. gene has no effect, then the gene is not important for the cell... For an engineer working on safety, the conclusion should be the exact opposite.



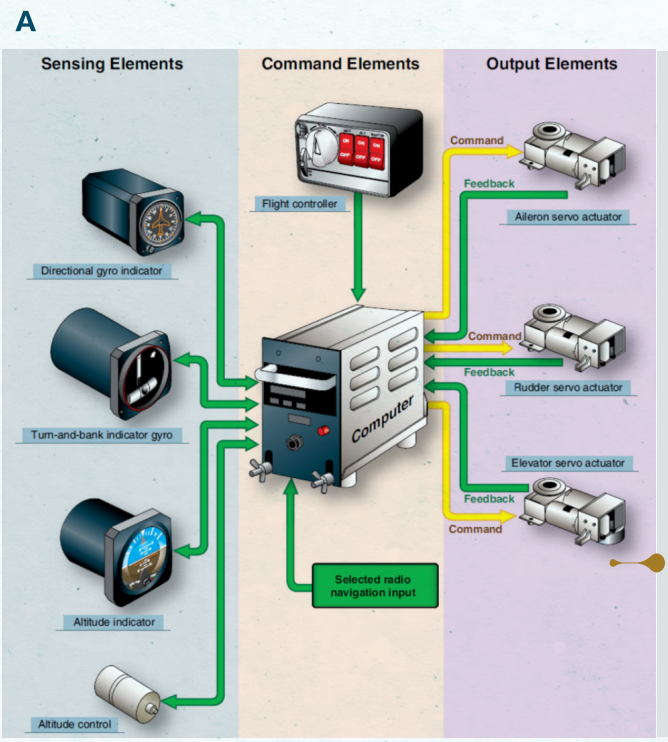
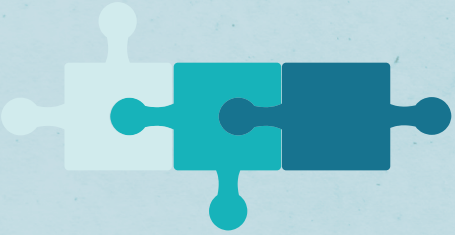


Fig. 4 As an autopilot, living systems use feedback loops to self-regulate. A) Components of a flight control loop in an autopilot. B) Representation of a cell from an engineering prism.



CONCLUSION

What are the common features between a cell and an aircraft autopilot? If one has studied enough both systems, they will be quite surprised by their similarities. The direct consequence is that you can borrow concepts and tools from engineering to unravel and analyze cell behaviors. Basics concepts like internal state, memory, time integration or feedback loops are essential to formalize and understand the behavior of autonomous systems like cells or autopilots. Cybernetics already tried to make such parallels during the last century. Today, with the complexification of engineered systems and our deeper comprehension of biological systems we should (re)consider common formalism to master those systems, so that biologists can, finally, fix a radio.

WHY DO WE OBSERVE SIMILARITIES BETWEEN BIOLOGICAL & ENGINEERED STRUCTURES?

Is that a bias interpretation from the engineer's point of view? The concept of “Convergent Evolution” [16] basically supposes that recurrent patterns found in life diversity are selected by external constraints from the environment. For a set of constraints, nature tends to select the same structure. Transposing this concept to autonomous and highly critical systems, like a cell or an autopilot that is submitted to a similar set of constraints, we find common patterns of structure. No matter if the structure was produced by random mutations or human intelligence, selection solely depends on external constraints.

The famous writer and pilot Antoine de Saint-Exupéry intuited in his book “Wind, Sand and Stars” the evolution of aircraft shape based on a process of evolution and selection:

« Have you looked at a modern airplane? Have you followed from year to year the evolution of its lines? Have you ever thought, not only about the airplane but about whatever man builds, that all of man's industrial efforts, all his computations and calculations, all the nights spent over working draughts and blueprints, invariably culminate in the production of a thing whose sole and guiding principle is the ultimate principle of simplicity?
It is as if there were a natural law which ordained that to achieve this end, to refine the curve of a piece of furniture, or a ship's keel, or the fuselage of an airplane, until gradually it partakes of the elementary purity of the curve of a human breast or shoulder, there must be the experimentation of several generations of craftsmen. In anything at all, perfection is finally attained not when there is no longer anything to add, but when there is no longer anything to take away, when a body has been stripped down to its nakedness. »

REFERENCES

1. Fundamental interaction. Wikipedia. In https://en.wikipedia.org/wiki/Fundamental_interaction
2. Y Lazebnik (2002). Can a biologist fix a radio? -Or, what I learned while studying apoptosis. Cancer Cell. doi: 10.1016/s1535-6108(02)00133-2.
3. HS Williams (1922) Practical Radio, Funk and Wagnalls Co., New York, p. 149
4. KJ Åström and Richard M. Murray (2008). Feedback Systems: An Introduction for Scientists and Engineers.. Princeton University Press. ISBN 0691135762.
5. R Rodriguez (2021). Adaptive Control (Part II) —Modeling the X-15's Adaptive Flight Control System. How to build from scratch a Simulink model of the famous MH-96, the X-15's Adaptive Flight Control System. In https://medium.com/technology-hits/adaptive-control-part-ii-modeling-the-x-15s-adaptive-flight-control-system-d784819f48e4
6. A Richard et al., (2016). Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process.O. PLoS Biol. 2016 doi: 10.1371/journal.pbio.1002585.
7. U Herbach et al. (2017). Inferring gene regulatory networks from single-cell data: a mechanistic approach. BMC Syst Biol. doi: 10.1186/s12918-017-0487-0
8. A Bonnaffoux et al., (2019). WASABI: a dynamic iterative framework for gene regulatory network inference. BMC Bioinformatics. 2019 May 2;20(1):220. doi: 10.1186/s12859-019-2798-1
9. Curse of dimensionality in https://en.wikipedia.org/wiki/Curse_of_dimensionality
10. Cybernetics in https://en.wikipedia.org/wiki/Cybernetics#History
11. L Goentoro et al., (2009). The incoherent feedforward loop can provide fold-change detection in gene regulation, Mol Cell. doi: 10.1016/j.molcel.2009.11.018.
12. K Takeda et al., (2012). Incoherent feedforward control governs adaptation of activated Ras in a eukaryotic chemotaxis pathway. Sci Signal. doi: 10.1126/scisignal.2002413
13. C Sokolik et al., (2015). Transcription factor competition allows embryonic stem cells to distinguish authentic signals from noise. Cell Syst. doi: 10.1016/j.cels.2015.08.001.
14. MJ Brenner et al., (2018) · Synthetic Biology: Immunotherapy by Design. Annu Rev Biomed Eng. doi: 10.1146/annurev-bioeng-062117-121147.
15. JR Huh et al., (2004). Compensatory proliferation induced by cell death in the Drosophila wing disc requires activity of the apical cell death caspase Dronc in a nonapoptotic role. Curr Biol. doi: 10.1016/j.cub.2004.06.015.
16. Anurag A. Agrawal, (2017). Toward a Predictive Framework for Convergent Evolution: Integrating Natural History, Genetic Mechanisms, and Consequences for the Diversity of Life. vol. 190, supplement the american naturalist. doi: 10.1086/692111

Cover page Image Credit:
Evan Ingersoll & Gael McGill "Cellular Landscape Cross-Section Through A Eukaryotic Cell" In https://t.co/mVs9dmh5bH

A QUESTION?
AN INFORMATION?
AN IDEA TO SHARE?

e-mail us at
contact@vidiumsolutions.com

A TRANS-DISCIPLINARY EXPERTISE

We have a very talented and human international team, where everyone has their say , to contribute to the continuous improvement of our solutions. Each and every one of our members are PhDs with deep trans-disciplinary expertise in the areas of engineering, data- science, biology, genomics, virology, immunology, medicine. . . We are united towards a common goal : to shift the paradigm of drug discovery and to bring new horizons in the area of medicine and life science in a sustainable way.

LET US GUIDE YOU THROUGH THE " LIFE CONQUEST " .

VIDIUM
SYSTEMS BIOLOGY SOLUTIONS

ENS de Lyon - 46 allée d›Italie
69007 Lyon - France