

Towards a better understanding of causation in cell biology

Brash, D. E.¹, Bizzarri, M.², Briscoe, J.³, Stern, C. D.⁴, Grieneisen, V. A.⁵,
and Levin, M.^{6*}

1 Depts. of Therapeutic Radiology and Dermatology and Yale Cancer Center, Yale School of Medicine, 333 Cedar St., New Haven, CT 06520-8040, USA

2 University La Sapienza, Dept. of Experimental Medicine, via Scarpa 16, 00161 Rome, Italy

3 The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK

4 Dept. Cell & Developmental Biology, University College London, Gower Street, London WC1E 6BT, UK

5 School of Biosciences, Cardiff University, Museum Avenue, Cardiff CF10 3AX, UK

6 Allen Discovery Center at Tufts University, 200 Boston Avenue, Suite 4600, Medford, MA 02155, USA

* Email: michael.levin@tufts.edu

What does it mean to say that event X caused outcome Y in biology? Explaining the causal structure underlying the dynamic function of living systems is a central goal of biology. Transformative advances in regenerative medicine and synthetic bioengineering require efficient strategies for causing desired system-level outcomes. We present a perspective on why it is critical to move beyond the classical ‘necessary and sufficient’ approach to define biological causality.

[H1] From genes to processes in developmental biology

Genes that control the development of specific tissues have been identified and organized into pathways, which has increased our understanding of how tissues are formed and what goes wrong in disorders and diseases. As a consequence, biological processes tend to be described by genes and gene networks. The result is a gene-centric view with genes assigned fixed and specific functions within hierarchical mechanisms, in which master regulators are the drivers of a developmental process. But developmental biology is more than a catalogue of parts. Understanding mechanisms must include a description of the activity and the causal relationship between components. This description shifts the attention to morphogenic patterns and connections between components, providing conceptual insight into how and why processes occur. Such change in perspective requires an interdisciplinary mix of theory and experiment. In cases where this approach has been followed ¹, the picture that arises reveals more sophistication and subtlety than implied by simple hierarchical genetic wiring diagrams. Nonlinearity and multi-level feedbacks in even small systems can have unexpected consequences. Reasoning that follows a linear logic becomes inadequate because the distinction between cause and effect is lost, and the explanation for how a process occurs will require an understanding of how relationships change over time. Describing and understanding these dynamics is now the challenge facing twenty-first century developmental biologists.

Through next-generation sequencing, protein interactomes, metabolomics and quantitative genetics, modern biology has become obsessed with compiling lists. But it is difficult, if not impossible, to understand processes from lists and merely undertaking the

effort at multiple scales does not suffice. Many of the current approaches to studying biological events have turned us away from asking the key questions needed to understand a biological process. Shifting our attention to patterns, connections and processes is what is needed, but not simply to make different types of lists. Emphasis should be placed on designing experiments to test alternative explanations for the observed behaviors. The experiments should make predictions of how one or another model might fit with the observations, so that the experiment, along with its controls, can distinguish between different mechanisms. This 'critical experiment' approach is the opposite of making lists, which effectively increases the number of dimensions [dimensions to consider? to include in a model?]. This change in focus and experimental design, will largely use the "IF (x is TRUE) THEN (y) ELSE (z)" logic of computational approaches. The approach becomes iterative — with each round of analysis including more data (produced by consecutive experiments) as well as more knowledge to reformulate the question. Ideal models make counter-intuitive predictions that can be tested rigorously — a crucial aspect of distinguishing between possible explanations.

Biophysical properties as causes

The current Gene Regulatory Network (GRN) formalism is not well-suited to explain the influence of spatially distributed physical factors, most of which manifest as constraints, arising from multiple interactions and physical fields. A constraint specifies some limits on independent behaviors of a dynamical system and this can provide access to new states unavailable to the unconstrained system. A specific case is provided by studies performed on mammalian cells cultured in microgravity that undergo changes in shape morphology and functions, leading to the partitioning of cells into two different populations. Yet, by separately re-seeding these populations once more in microgravity, two distinct phenotypes emerge from each cell cluster³. This experiment demonstrates that a biological field 'deprived' of the gravity constraint does not provide a stable environment for differentiation, as it favors a never-ending transition among different phenotypes leading to a consequent change in the GRN state of activation. In other words, in the absence of physical constraints the environment cannot select between two different phenotypes due to the emergence of a 'permanent transition state', while the

presence of physical cues force the system into a specific differentiated state. Constraints, by providing the system with a deterministic output that would otherwise have been impossible to obtain, act thereby as organizing principles during cell differentiation.

If we look beyond the single cell level, regulative embryogenesis and regeneration exemplify the still poorly understood decision-making processes by which cells cooperate toward the dynamic maintenance and repair of complex 3-dimensional structures⁴. Salamanders regenerate entire limbs and, crucially, they stop precisely when a correct limb structure is complete. Tadpoles with highly abnormal faces nevertheless can become normal frogs, underscoring the fact that genomes encode not hardwired cell movements for metamorphosis but a system of large-scale remodeling that implements profound plasticity, reaching the species-specific *target* morphology despite drastic interventions. Remarkably, the setpoint of this anatomical homeostasis can be re-written, e.g. as trophic memory. In deer antler regeneration, ectopic tines will form at sites of prior years' injuries in an antler rack replacing one that has already been shed; in planarian axial patterning, a transient stimulus can make permanent lines of genetically-normal 2-headed worms. An important component of such control systems appears to be biophysical, with instructive information encoded by bioelectric state dynamics distributed across tissues. The anatomical pattern that defines the target (and the stop condition) for regeneration can be permanently altered, for example by transiently modulating the state of bioelectric circuits. This strategy is already being used in model systems for repairing birth defects and inducing complex regeneration. Expanding comfortable paradigms of master regulator genes and cell-level epigenetics to include a physiological layer that enables robust computation and pattern memory will be essential to improving the understanding and control of the adaptive remodeling of growth and form.

[H1] Cause and constraint, physics and biology

The billiard ball model of causality has lured biologists into using knock-down and overexpression approaches to test linear, unidirectional models of biological systems that are not actually built this way.[we should change this sentence] Linearity is broken by branching pathways and unidirectionality is broken by feedback, which can occur on different levels of biological organization (molecular, cellular, tissue-, organ- and

organism-level). Models of single-level interactions are broken by biological hierarchy. An archetype is the Chladni plate. If we sprinkle a flat plate with sand grains and set it to vibrate, distinctive patterns emerge, which will differ depending on shape or size of the plate, with different patterns emerging as a result of vibration frequency changes or local constraints, such as our thumb placed on the plate's corner. The pattern results from collisions of sand grains governed by Newton's Laws. The precise collisions might be computed, but are in fact irrelevant, as the collisions will be different tomorrow. However, the patterns are reproducible independently of these alterations, remaining consistent with the external constraints such as the plate's size and shape, and the frequency of vibration. So what is meaningful for understanding the patterns is not to seek a description of the links between sand grains but the rules that link the higher-level patterns to constraints — a new form of causality focusing on determinants rather than instigators. A key first step is to identify the constraints. For example ... [a sentence about the gravity example?]. Three useful refinements of causality in physics and engineering can be imported into biology: A shift from causes acting at a single site in an organism and instigating changes in a linear pathway, to looking at the behavior of the entire system, including the constraints that shape it. A shift from studying molecular events to discrete pattern states, "eigenisms" analogous to the eigenstates of quantum mechanics that can only exist in a small number of distinctly different states. A shift from medicines that briefly control a single target to treatments that put constraints on many parts of the organism, sustained over time.

Furthermore, to these principles, biology adds an inevitable duality of causation. Ludwig von Bertalanffy described the state of biology in 1928 as "*dominated by the 'mechanistic' world view... that is, it (biology) tried to model itself according to the pattern set by physics.*".

Biological causation goes beyond asking how a system of proteins establishes heterogeneous patterning, and seeks to explain what the biological function is. For example, stating that small GTPases are activated by GAPs and inactivated by GEFs, and that GTPase activity regulates downstream events, is like stating that A leads to B. However, defining the function of these interactions (providing intrinsic polarity

information to the cell) requires us to operate on a different conceptual level. This is in stark contrast with physics, in which a causal understanding of physical phenomena is (generally) devoid of function. By comparing species that have evolved multicellularity independently, such as in the animal and plant kingdoms, we are now able to search for fundamental mechanisms adopted by life, by identifying conserved and modified functions <?> rather than conserved components (such as protein structures, DNA, etc). Such comparative approaches, using understanding of system processes rather than system components, gives insight into how tissue polarity is established and why it has done so in a particular manner. Thus, causation in biology should define how *and* why a process occurs in a certain manner. Notwithstanding the qualitative challenge of defining such biological causation, the quantitative approaches of mathematical modelling and experimentation are required ⁵ to test quantitative constraints needed to appropriately direct a biological process.

[H1] Conclusions and outlook

Despite biologists' general awareness of redundancy and homeostatic control circuits, we still largely do not understand the corrective, self-organizing processes that reliably reach complex, systems-level patterning goals. This insight is key in contexts where the outcome is an emergent result of coupled, parallel processes, and it will be increasingly important for knowing when machine learning strategies could be an appropriate approach to extract actionable, efficient control policies from the ever-increasing deluge of data. Importantly, extensive work in fields such as philosophy, physics and engineering has identified profound problems with respect to naïve models of causation ⁶, which still pervade experimental biology. Advances in mathematical approaches that extract causal control structures from biological data ⁷ and an increased awareness of the presence and importance of variability, stochasticity, heterogeneity, and noise are revolutionizing the search for effective causes of complex biological states and processes. An appreciation of new kinds of instructive influences, from network science and physics, must be coupled with hypothesis testing, focused on models of functional processes (not necessarily molecular components). Developing methodology for the rigorous identification and efficient control of causal processes in complex biological

systems is very much a nascent field that has not been widely integrated with molecular approaches to growth and form. At stake is not only perceptive review of manuscripts and grants (in which distinguishing between causes and epiphenomena is a key goal), but the development of efficacious next-generation interventions in regenerative, cancer and synthetic biology.

Competing interests

The authors declare no competing interests.

References

- 1 Zagorski, M. *et al.* Decoding of position in the developing neural tube from antiparallel morphogen gradients. *Science* **356**, 1379-1383, doi:10.1126/science.aam5887 (2017).
- 3 Bizzarri, M., Masiello, M. G., Giuliani, A. & Cucina, A. Gravity Constraints Drive Biological Systems Toward Specific Organization Patterns: Commitment of cell specification is constrained by physical cues. *Bioessays* **40**, doi:10.1002/bies.201700138 (2018).
- 4 Pezzulo, G. & Levin, M. Re-membering the body: applications of computational neuroscience to the top-down control of regeneration of limbs and other complex organs. *Integr Biol (Camb)* **7**, 1487-1517, doi:10.1039/c5ib00221d (2015).
- 5 Grieneisen, V. A., Scheres, B., Hogeweg, P. & AF, M. M. Morphogengineering roots: comparing mechanisms of morphogen gradient formation. *BMC Syst Biol* **6**, 37, doi:10.1186/1752-0509-6-37 (2012).
- 6 Hoel, E. P., Albantakis, L. & Tononi, G. Quantifying causal emergence shows that macro can beat micro. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 19790-19795, doi:10.1073/pnas.1314922110 (2013).
- 7 Tsuchiya, M., Giuliani, A., Hashimoto, M., Erenpreisa, J. & Yoshikawa, K. Self-Organizing Global Gene Expression Regulated through Criticality: Mechanism of the Cell-Fate Change. *PLoS One* **11**, e0167912, doi:10.1371/journal.pone.0167912 (2016).