IOP Publishing

Ahead of the Curve
Hidden breakthroughs in the biosciences
Michael Levin and Dany Spencer Adams

Chapter 5

Mathematics and modeling

Case Study 14: Top-Down Causation: Not All The Work Is Done By Molecules

doi:10.1088/978-0-7503-1326-1ch5

Perspective on *Quantifying causal emergence shows that macro can beat micro*

Sara Imari Walker,
Arizona State University

More is Not Just Different, its Causal

The reductionist paradigm in science has worked remarkably well for almost four centuries, ever since Newton consolidated much of our understanding of motion into a few concise equations. In the ensuing centuries, reductionism has given us insights into the structure of the atom, the fundamental forces of nature, and the inner workings of the cell among many others. Nonetheless there remain hints that reductionism qua reductionism cannot be the entire story. Fundamental components of reality—including life and mind—are not easily reconciled with the idea that all of reality may be reduced to description solely in terms of those fundamental laws that operate at its most microscopic layers. In the words of the Nobel Laureate Philip Anderson in his famous essay *More is Different* [1]:

> *'The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe.'*

As we move up in scale, the universe appears to become more and more complex, and it seems clear that as complexity increases entirely new properties emerge. For example, *How is it that the collection of atoms in a human brain is organized to give rise to thoughts, feelings and urges?* Perhaps more mysterious is that our thoughts appear to matter to the world. A simple thought experiment (no pun intended) of intending to lift your left arm and then executing that action provides the evidence. There is intense debate about whether this kind of example represents a conscious or subconscious action (it may not matter for this debate), and specifically whether one can really consider the action to be *caused* by a thought in your brain. A thought is an informational abstraction—so the debate really centers on whether *information* can have causal power. If all of reality can be described by mechanical actions at the micro-level only, there is no 'room at the top' for your thoughts to be causal. Stated succinctly, the challenge we are faced with is to determine what is doing the causal work—*is it you or your atoms?*

Settling the debate requires quantifying causal emergence, such that we can identify where the 'causal work' actually lies. There has been a long-held assumption that the bottom-up picture of reality excludes higher-levels from playing a role. This is refuted by the work of Hoel *et al* who explicitly quantify *cause* and *effect* information and show that more causal power can exist at macroscopic levels in some systems, depending on how they are networked together [2]. The key result is that even though there may exist a complete lower-level description of a physical system, higher levels can still be causal above and beyond the causal work of lower levels. Emergence and reductionism are not at odds. In fact, reductionism is what makes emergence possible [3]: if we could not reduce reality to the study of a few

component building blocks, we would not be able to describe how those building blocks come together to create more complex 'higher-levels'.

This is significant since there are many examples across the sciences where 'top-down' causal effects are essential to building a coherent explanatory narrative [4]. Our conscious experience is only the most blatant example. Others come from quantum physics, thermodynamics, complex systems science and biology. For the latter, a causal role for macroscopic entities may even play a critical role in defining its life's origins [5]. In short, these same issues emerge (pun!) anywhere the term 'emergence' is cast about. We need a theory of causal emergence to understand the hierarchical organization of our universe, and why 'higher-levels' not only appear so much richer and more complex than the lower-levels they emerge from, but why they seem to matter at all.

This story is just beginning. While the idea of causal emergence has been demonstrated as a proof of concept, it is unclear if the formalism as presented will be the ultimate one science settles on. Important open questions remain. For example, causal emergence occurs in networks with particular architectures that are highly non-linear and include feedback loops and irreversible logic operations. Clearly these kinds of structures exist at macroscopic levels in the human brain, but how are these compatible with our lowest level descriptions of reality in physics, which are cast in terms of initial states and local, reversible laws of motion? Can causal emergence still occur if we consider the deepest layers of reality and do not already start from a macroscopic description? There also remains the open question of whether the macro could actually *intervene* on the micro (and not just beat it), such that top-down causation might be said to truly occur. These questions represent the next frontier for 21st century science to explain the emergent layers of reality as we experience them.

References

[1] Anderson P W 1972 More is Different *Science, New Series* Vol. 177 No. 4047 393–6

[2] Hoel E P, Albantakis L and Tononi G 2013 Quantifying causal emergence shows that macro can beat micro *Proc. Natl Acad. Sci.* **110** 19790–5

[3] Smith E and Morowitz H J 2016 *The Origin and Nature of Life on Earth: The Emergence of the Fourth Geosphere* (Cambridge University Press)

[4] Ellis G F R "Top-down causation and emergence: some comments on mechanisms." *Interface Focus* 2011 rsfs20110062

[5] Walker S I and Davies P C W 2013 The Algorithmic Origins of Life *J. Roy. Soc. Interface* **6** 20120860

Quantifying causal emergence shows that macro can beat micro

Erik P. Hoel, Larissa Albantakis, and Giulio Tononi[1]

Department of Psychiatry, University of Wisconsin, Madison, WI 53719

Edited by Michael S. Gazzaniga, University of California, Santa Barbara, CA, and approved October 22, 2013 (received for review August 6, 2013)

Causal interactions within complex systems can be analyzed at multiple spatial and temporal scales. For example, the brain can be analyzed at the level of neurons, neuronal groups, and areas, over tens, hundreds, or thousands of milliseconds. It is widely assumed that, once a micro level is fixed, macro levels are fixed too, a relation called supervenience. It is also assumed that, although macro descriptions may be convenient, only the micro level is causally complete, because it includes every detail, thus leaving no room for causation at the macro level. However, this assumption can only be evaluated under a proper measure of causation. Here, we use a measure [effective information (*EI*)] that depends on both the effectiveness of a system's mechanisms and the size of its state space: *EI* is higher the more the mechanisms constrain the system's possible past and future states. By measuring *EI* at micro and macro levels in simple systems whose micro mechanisms are fixed, we show that for certain causal architectures *EI* can peak at a macro level in space and/or time. This happens when coarse-grained macro mechanisms are more effective (more deterministic and/or less degenerate) than the underlying micro mechanisms, to an extent that overcomes the smaller state space. Thus, although the macro level supervenes upon the micro, it can supersede it causally, leading to genuine causal emergence—the gain in *EI* when moving from a micro to a macro level of analysis.

In science, it is usually assumed that, the better one can characterize the detailed causal mechanisms of a complex system, the more one can understand how the system works. At times, it may be convenient to resort to a "macro"-level description, either because not all of the "micro"-level data are available, or because a rough model may suffice for one's purposes. However, a complete understanding of how a system functions, and the ability to predict its behavior precisely, would seem to require the full knowledge of causal interactions at the micro level. For example, the brain can be characterized at a macro scale of brain regions and pathways, a meso scale of local populations of neurons such as minicolumns and their connectivity, and a micro scale of neurons and their synapses (1). With the goal

[1] To whom correspondence should be addressed. E-mail: gtononi@wisc.edu.

Author contributions: E.P.H., L.A., and G.T. designed research; E.P.H. and L.A. performed research; E.P.H., L.A., and G.T. analyzed data; and E.P.H., L.A., and G.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1314922110/-/DCSupplemental.

of a complete mechanistic understanding of the brain, ambitious programs have been launched with the aim of modeling its micro scale (2).

The reductionist approach common in science has been successful not only in practice, but has also been supported by strong theoretical arguments. The chief argument starts from the intuitive notion that, when the properties of micro-level physical mechanisms of a system are fixed, so are the properties of all its macro levels—a relation called "supervenience" (3). In turn, this relation is usually taken to imply that the micro mechanisms do all of the causal work, i.e., the micro level is causally complete. This leaves no room for any causal contribution at the macro level; otherwise, there would be "multiple causation" (4). This "causal exclusion" argument is often applied to argue against the possibility for mental causation above and beyond physical causation (5), but it can be extended to all cases of supervenience, including the hierarchy of the sciences (6).

Some have nevertheless argued for the possibility that genuine emergence can occur. Purported examples go all of the way from the behavior of flocks of organisms (7) to that of ant colonies (8), brains (9), and human societies (10). Unfortunately, it remains unclear what would qualify some systems as truly emergent and others as reducible to their micro elements. Also, most arguments in favor of emergence have been qualitative (11). A convincing case for emergence must demonstrate that higher levels can be causal above and beyond lower levels ["causal emergence" (*CE*)]. So far, the few attempts to characterize emergence quantitatively (12) have not been based on causal models.

Here, we make use of simple simulated systems, including neural- like ones, to show quantitatively that the macro level can causally supersede the micro level, i.e., causal emergence can occur. We do so by perturbing each system through its entire repertoire of possible causal states ("counterfactuals," in the general sense of alternative possibilities) and evaluating the resulting effects using "effective information" (*EI*) (13). *EI* is a general measure for causal interactions because it uses perturbations to capture the effectiveness/selectivity of the mechanisms of a system in relation to the size of its state space. As will be pointed out, *EI* is maximal for systems that are deterministic and not degenerate, and decreases with noise (causal divergence) and/or degeneracy (causal convergence).

For each system, we completely characterize the causal mechanisms at the micro level, fixing what can happen at any macro level (supervenience). Macro levels are defined by coarse graining the micro elements in space and/or time, and this mapping defines the repertoire of possible causes and effects at each level. By comparing *EI* at different levels, we show that, depending on how a system is organized, causal interactions can peak at a macro rather than at a micro spatiotemporal scale. Thus, the macro may be causally superior to the micro even though it supervenes upon it. Evaluating the changes in *EI* that arise from coarse or fine graining a system provides a straightforward way of quantifying both emergence and reduction.

Theory

In what follows, we consider discrete systems S of connected binary micro elements that implement logical functions (mechanisms) over their inputs. We first introduce a

state-dependent measure of causation, the "cause" and "effect information" of a single system state s_0, before we describe the state-independent *EI* of the system *S*.

Significance

Properly characterizing emergence requires a causal approach. Here, we construct causal models of simple systems at micro and macro spatiotemporal scales and measure their causal effectiveness using a general measure of causation [effective information (*EI*)]. *EI* is dependent on the size of the system's state space and reflects key properties of causation (selectivity, determinism, and degeneracy). Although in the example systems the macro mechanisms are completely specified by their underlying micro mechanisms, *EI* can nevertheless peak at a macro spatiotemporal scale. This approach leads to a straightforward way of quantifying causal emergence as the supersedence of a macro causal model over a micro one.

State-Dependent Causal Analysis. The micro mechanisms of *S* specify its state-to-state transition probability matrix (TPM) at a micro time step *t*. Building upon the perturbational framework of causal analysis developed by Judea Pearl (14; see also ref. 18), the TPM can be obtained by perturbing *S* at t_0 (13) into all possible *n* initial states with equal probability $1/n$ [$\mathrm{do}(S = s_i)\ \forall i \in \ldots n$]. Perturbing the system in this way corresponds to the unconstrained repertoire (probability distribution) of possible causes U^C and determines the probability of the resulting states at t_{+1}, corresponding to the unconstrained repertoire of possible effects U^E. Although U^C is thus identical to the uniform distribution U [with $p(s) = 1/n, \forall s \in S$], U^E is typically not uniform. A current system state $S = s_0$ is associated with the probability distribution of past states that could have caused it ("cause repertoire $S_P|s_0$," obtained by Bayes' rule), and the probability distribution of future states that could be its effects ("effect repertoire $S_F|s_0$"). A system's mechanisms and current state thus constrain both the repertoire of possible causes U^C and that of possible effects U^E. An informational measure of the causal interactions in the system (15) can then be defined as the difference [here Kullback–Leibler divergence (D_{KL}) (16)] between the constrained and unconstrained distributions:

$$\begin{aligned}\text{Causeinformation}(s_0) &= D_{\mathrm{KL}}((S_P|S_0),\ U^C),\\ \text{Effect information}(s_0) &= D_{\mathrm{KL}}((S_F|s_0),\ U^E).\end{aligned}$$

Cause/effect information depends on two properties: (*i*) the size of the system's state space (repertoire of alternatives), because both are bounded by $\log_2(\mathrm{n})$; (*ii*) the effectiveness of the system's mechanisms in specifying past and future states. To isolate effectiveness from size, we define the following normalized coefficients:

$$\text{Cause coefficient}(s_0) = \frac{\text{Cause Information}(s_0)}{\log_2(n)},$$
$$\text{Effect coefficient}(s_0) = \frac{\text{Effect Information}(s_0)}{\log_2(n)}.$$

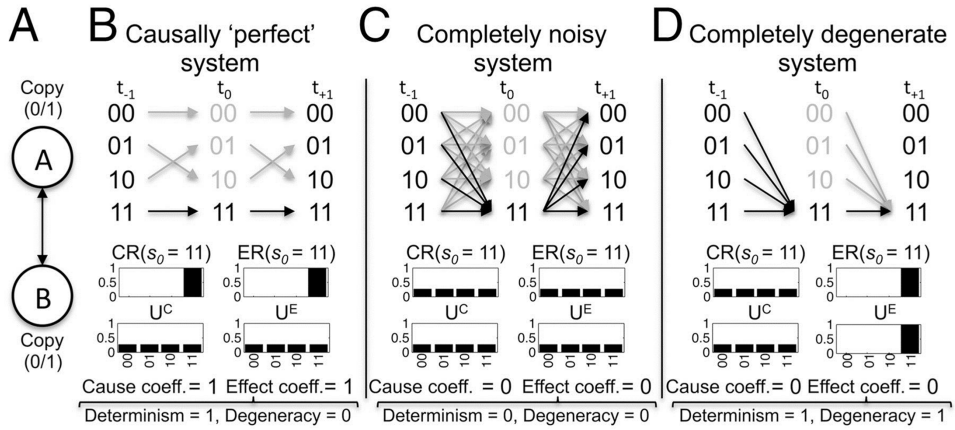


Fig. 1. Cause and effect coefficients in example systems with different causal architectures. (*A*) The systems consist of two interconnected binary COPY gates with possible states 0 and 1. (*B*) A causally perfect system, in which each state has one cause and one effect. Thus, s_0 = [11] has a cause and effect coefficient (coef.) of 1. Moreover, there is no divergence (determinism coef. = 1) and no convergence (degeneracy coef. = 0). (*C* and *D*) In both the completely indeterministic and completely degenerate systems, state s_0 = [11] is completely insufficient to specify past system states and completely unnecessary to specify future states (cause and effect coefficient = 0). Note that the degeneracy coef. is 0 in the completely noisy system, because all convergence is due to noise alone.

The "cause coefficient" describes to what extent a state is sufficient to specify its past causes, and the "effect coefficient" indicates how necessary a state is to specify its future effects (Fig. 1*B*). In turn, the effect coefficient itself is a function of two terms, "determinism" and "degeneracy" (see *Effect Coefficient and Effectiveness (Eff) Expressed as Determinism and Degeneracy* for derivation):

$$\begin{aligned}\text{Effect coefficient} &= \text{Determinism coefficeint}\,(s_0) \;-\; \text{Degeneracy coefficient}\,(s_0)\\ &= \frac{1}{\log_2(n)} \sum_{S_F \in U^E} p(S_F|S_0)\log_2(n \cdot p(S_F|S_0))\\ &\quad - \frac{1}{\log_2(n)} \sum_{S_F \in U^E} p(S_F|S_0)\log_2(n \cdot p(S_F|S_0)).\end{aligned}$$

The determinism coef. is the difference D_{KL} (($S_F|s_0$), U) between the effect repertoire and the uniform distribution (U) of system states, divided by log2(n), and measures how deterministically (reliably) s_0 leads to the future state of the system: it is "1" (complete determinism) when the current state leads to a single future state with probability $p = 1$, and is "0" (complete in-determinism or noise) if it could be followed by every future state with $p = 1/n$. The degeneracy coef. measures to what degree there is deterministic convergence (not due to noise) from other states onto the future states specified by s_0. In broad terms, degeneracy refers to multiple ways of deterministically achieving the same effect or function (17, 18). The degeneracy coef. is 1 (complete degeneracy) when s_0 specifies the same future state as all other states, and 0 when s_0 specifies a unique future state (no degeneracy).

Both cause and effect coefficients are minimal (0) in a completely noisy or completely degenerate system (Fig. 1 *C* and *D*) and maximal (1) in a deterministic,

nondegenerate system (*Bounds of Cause and Effect Coefficients and Effectiveness Eff (S)*. The contribution of a single state to the system's determinism and degeneracy are best demonstrated by decomposing the effect coefficient. Although the cause coefficient also reflects the degeneracy and determinism of the system, it is not subdivided further here.

State-Independent Causal Analysis. A state-independent informational measure of a system's causal architecture can be obtained by taking the expected value of cause or effect information over all system states, a quantity called effective information (*EI*):

$$EI(s) = \langle \text{Cause Information}(s_0) \rangle = \sum_{s_0 \in U^E} p(s_0) D_{\mathrm{KL}}((S_P|s_0),\ U^C)$$

$$= \langle \text{Effect Information}(s_0) \rangle = \frac{1}{n} \sum_{s_0 \in U^E} D_{\mathrm{KL}}((S_F|s_0),\ U^E).$$

The two terms are identical, because the system is assumed to be time invariant ($\langle t_{-1} \to t_0 \rangle = \langle t_0 \to t_{+1} \rangle$), and cause and effect information are related via Bayes' rule. *EI* is also the mutual information *(MI)* between all possible causes and their effects, $\mathrm{MI}(\mathrm{U}^{\mathrm{C}};\mathrm{U}^{E})$ (Effective Information EI(S) Expressed in Terms of Cause and Effect Information and Mutual Information MI).

As a measure of causation, EI captures how effectively (deterministically and uniquely) causes produce effects in the system, and how selectively causes can be identified from effects. As with the state-dependent measures, the effectiveness (Eff) of the causal interactions within a system can be captured by normalizing EI by the system's size: $Eff\ (S) = \mathrm{EI}(S)/\log_2(n)$. Also as in the state-independent case, effectiveness can be split into two components, determinism and degeneracy:

$$\begin{aligned} Eff(S) &= \langle \text{Determinism coefficient}(s_0) \rangle \\ &\quad - \langle \text{Degeneracy coefficient}(s_0) \rangle \\ &= \left\langle D_{\mathrm{KL}}((S_F|s_0),\ U) \right\rangle / \log_2(n) - D_{\mathrm{KL}}(U^E|U)/\log_2(n). \end{aligned}$$

Thus, $Eff(\mathrm{S}) = 1$ if *EI* is maximal for a given system size, and decreases with indeterminism (divergence due to noise) or degeneracy (deterministic convergence), with $Eff(\mathrm{S}) = 0$ for completely noisy or degenerate systems (Fig. 1 C and D). In a system with perfect effectiveness (Fig. 1*B*), each cause has a unique effect, and each effect has a unique cause. Thus, such a system [where $Eff(S) = 1$] is perfectly retrodictive/predictive, in the sense that not only the unique future trajectory, but also the unique past trajectory of all states can be deduced from the TPM (complete causal reversibility).

Levels of Analysis. A finite, discrete system S can be considered at various levels, from the most fine-grained micro causal model S_m through various coarse-grained causal models S_M. All macro levels S_M are assumed to be "supervenient" on the micro level S_m: given the micro elements of S_m and the causal relationships between them, all other members of $\{\mathbf{S}\}$—the set of all possible causal models of system S—are fixed as well (19). Although S_m fixes S_M, any S_M may be fixed by a number of different lower level descriptions, a property known as "multiple realizability" (20).

Groupings. Micro elements are binary and labeled by Latin letters {A, B, C ...}, macro elements by Greek letters {α, β, γ …}. Micro states are labeled {1, 0} and macro states {"on," "bursting," "quiet"…). Micro elements can be grouped into macro elements spatially, temporally, or both. Micro states are grouped into macro states through a mapping $M : S_m \rightarrow S_M$. The mapping must be exhaustive and disjunctive over micro elements (all of the states of one micro element must be mapped to the states of the same macro element; note that a macro element can consist of a single micro element as long as the state space of the system is reduced). Moreover, the mapping must be such that no micro-level information is available at the macro level (the identity of the micro elements grouped into a macro element is lost). For example, the grouping of the four states of two micro elements into the two states of one macro element as [[00, 01, 10] = off, [11] = on] is permitted, whereas the grouping [[00, 01], [10, 11]] is not, because distinguishing 01 from 10 requires knowing the identity of the micro elements.

Level-Specific Perturbations. Causal analysis at the micro level S_m, requires setting S into all possible micro states with equal probability (i.e., testing all micro alternatives) and determining the resulting effects. When moving to a macro level S_M, S must similarly be set into all possible macro states with equal probability (i.e., testing all macro alternatives). To causally assess any macro state, then, one must set S into all of the n_{micro} micro states $\{s_m\}$ that are grouped into the corresponding macro state s_M, and average over the effects. This is done using a "macro perturbation": $do(S_M = s_M) = \frac{1}{n_{micro}} \sum_{s_{m,\,i} \in S_M} do(S_M = s_{m,\,i})$. Using such macro perturbations, one can obtain cause/effect information and EI for every coarse grain of S_m. EI at each macro level is then equivalent to the MI between the set of macro causes and their macro effects.

Causal Emergence/Reduction. Finally, by assessing $EI(S)$ over all coarse grains of S_m, one can ask at which level of $\{\mathbf{S}\}$ causation reaches a maximum. This provides an analytical definition of causal emergence, expressed in bits: $CE = \underline{\mathrm{EI}}(S_M) - EI(S_m)$.

Thus, if $EI(S)$ is maximal for a macro-level S_M rather than the micro-level S_m, then $CE > 0$ and causal emergence occurs. If for every macro-level $CE < 0$, causal reduction holds. Although the focus here is on emergence/reduction relative to the micro-level S_m, the above measure can of course be used to compare different macro levels.

As mentioned above, $EI(S)$ depends on both the size of the system's repertoire of states and on the effectiveness of its mechanisms. When moving from one system level to another, both terms change as the state space becomes smaller or larger, and the individual states become more or less selective with respect to the past, and more or less determined or degenerate with respect to the future. The respective informational contributions of repertoire size and effectiveness to $\Delta EI(S)$ can be expressed separately as follows: $\Delta I_{Eff} = (Eff\,(S_M) - Eff(S_m)) \cdot log_2(n_M)$, $\Delta I_{Size} = Eff(S_m) \cdot (log_2(n_M) - log_2(n_m))$, where $n_{m/M}$ is the state repertoire size of $S_{m/M}$. It follows that $\Delta EI = \Delta I_{Eff} + \Delta I_{Size} = CE$. A positive ΔI_{Eff} can thus be due to the macro reducing the degeneracy of the micro level, increasing the determinism of the micro level, or

both. Notably, coarse graining the micro-level S_m into macro-level S_M implies that ΔI_{Size} is always negative. Hence, for causal emergence to occur [$EI(S_M) > EI(S_m)$], the increase in effectiveness ΔI_{Eff} must outweigh the decrease in ΔI_{Size}.

Results

Causal analysis was performed across all coarse grains of a system [only the S_M with maximal $EI(S)$ is shown in the figures] with a custom-made Python program. Data plots were created using MATLAB. Below, we consider examples of spatial, temporal, and spatiotemporal emergence (see Fig. S1 for an example of spatial reduction).

Spatial Causal Emergence. As a proof-of-principle example, consider a system of four binary elements S_m = {ABCD} (Fig. 2A). Each micro mechanism is an AND-gate (two inputs) operating over some intrinsic noise. The 16 × 16 S_m TPM was constructed by setting the system into all possible micro states from [0000] to [1111] with equal probability (Fig. 2*B*). At the micro level S_m, effective information $EI(S) = 1.15$ bits, out of maximally 4 bits, with effectiveness $Eff(S_m) = 0.29$. The macro level S_M (Fig. 2*D*), composed of two elements {α, β}, each with states {"on," "off"}, is a coarse graining of S_m as defined by the mapping $\boldsymbol{M}$ in Fig. 2*C*. The 4 × 4 S_M TPM was obtained by setting the system into all possible macro states from [off, off] to [on, on] with equal probability (Fig. 2*E*). For the macro level, $EI(S_M) = 1.55$ bits, higher than $EI(S_m) = 1.15$ bits. Thus, $CE(S) = 0.40$ bits, demonstrating that in

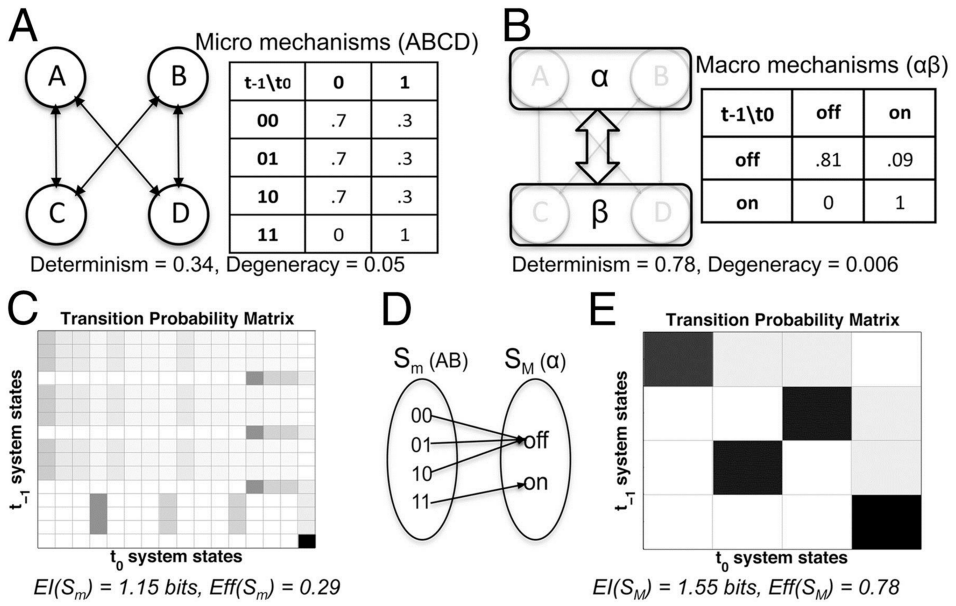


Fig. 2. Spatial causal emergence (counteracting indeterminism). (*A*) The micro level S_m of system S is composed of identical noisy micro mechanisms. (*B*) The micro TPM. (*C*) A macro causal level S_M and its TPM are defined by the mapping $\boldsymbol{M}$ (shown for AB to α, CD to β is symmetric). (*D*) S_M and its macro mechanisms. (*E*) By reducing indeterminism and increasing effectiveness *Eff*, the macro beats the micro in terms of *EI* despite the reduced repertoire size ($CE = 0.40$ bits).

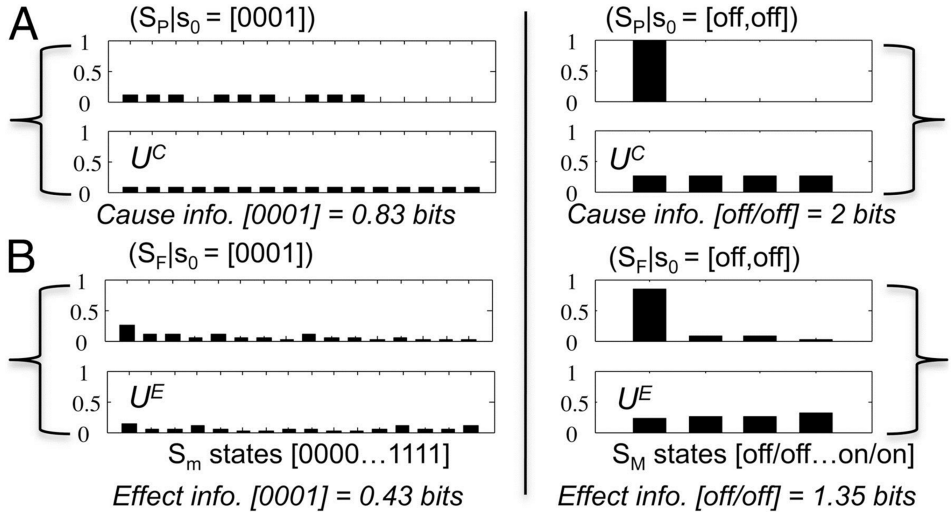


Fig. 3. State-dependent cause/effect information. (*A*) The cause information of S_m in micro state {ABCD} = [0001] is calculated as the difference (D_{KL}) between the cause repertoire of state [0001] and the unconstrained micro repertoire U^C (*Left*). The cause information of S_M in the supervening macro state {αβ} = [off/off] (*Right*) is the difference (D_{KL}), between the cause repertoire of [off/off] and the unconstrained macro repertoire U^C. (*B*) Effect information. The higher cause and effect information at the macro level is due to an increase in determinism and decrease in degeneracy, reflecting higher selectivity.

this case the macro S_M beats the micro S_m and constitutes the optimal causal model of system S. This is because the TPM for S_M is much closer to perfect effectiveness [$Eff(SM) = 0.78$] and the increase in effectiveness gained by grouping $\Delta I_{Eff} = 0.97$ bits outweighs the loss in size $\Delta I_{Size} = -0.57$ bits. In this example, the gain in effectiveness ΔI_{Eff} at the macro level comes primarily (91%) from counteracting noise [determinism coef. (Sm) = 0.34; (S_M) = 0.78] and less so (9%) from reducing degeneracy [degeneracy coef. (S_m) = 0.05; (S_M) = 0.006].

The higher effectiveness of the macro level is also evident comparing S_m and S_M in a state-dependent manner. As an example, the cause/effect distributions for S_m in state {ABCD} = [0001] are compared with the corresponding S_M state {αβ} = [off, off] in Fig. 3. Comparing the cause/effect distributions of S_m = [0001] against the unconstrained repertoires (using D_{KL}) yields 0.83 bits of cause information and 0.43 bits of effect information. For the macro S_M, cause information is 2 bits and effect information 1.35 bits. The macro beats the micro because {αβ} = [off, off] is both more selective and more reliable than {ABCD} = [0001].

Causal emergence may arise not only from macro gains in determinism (as above), but also from reducing degeneracy. In Fig. 4, micro elements A-F are deterministic AND gates connected in away that ensures high degeneracy (Fig. *4A*, determinism coef. = 1; degeneracy coef. = 0.6), resulting in $Eff(S_m) = 0.4$ and $EI(S_m) = 2.43$ bits (Fig. 4*C*). The optimal macro groups the six micro AND gates into three macro COPY gates (αβγ) (Fig. 4*B*). Both macro and micro are deterministic, but by eliminating degeneracy $\Delta I_{Eff} = 1.79$ bits $> -\Delta I_{Size} = 1.22$ bits. As a result, Eff (S_M) = 1, $EI(S_M) = 3$ bits, and the macro emerges over the micro ($CE = 0.57$ bits).

Temporal Causal Emergence. The same principles allowing for emergence through spatial groupings hold for temporal groupings, which coarse grain micro time steps (t_x) into macro time steps (T_x). The example in Fig. 5 shows micro elements that, upon receiving an input "burst" of two spikes, respond with an output burst of two spikes. Thus, elements implement second-order Markov mechanisms over both inputs and outputs (Fig. 5*A*). Fig. 5*B* shows that causal interactions assessed over one micro time step are weak [$EI(S_m) = 0.16$ bits; $Eff(S_m) = 0.03$] because they fail to capture the second-order mechanisms. By contrast, causal analysis over two micro time steps (Fig. 5*C*) gives $EI = 1.38$ bits and $Eff(S_m) = 0.34$. The temporal grouping of micro into macro states $\alpha = \{A_t, A_{t+1}\}$ and $\beta = \{B_t, B_{t+1}\}$ (Fig. 5*D*) is analogous to the spatial grouping in Fig. 2: {00, 01, 10} = {off} and {11} = {on}. Over macro time steps, the system becomes fully deterministic and nondegenerate, $EI(S_M) = 2$ bits, $Eff(S_M) = 1$, and $CE(S) = 0.62$ bits (Fig. 5 *E* and F).

Spatiotemporal Causal Emergence. In general, emergence may occur simultaneously over space and time (Fig. 6). As in Fig.5, the nine neural-like micro elements in Fig. 6*A* are second-order Markov mechanisms, integrating inputs and outputs over two micro time steps, $t_{-2}\, t_{-1}$, and $t_0\, t_{+1}$, respectively [compare to longer time constants of NMDA receptors (21)]. Moreover, in the examples above, the micro elements within a macro element were not connected and were causally equivalent. To demonstrate that this is not a requisite for causal emergence, in Fig. 6, the micro elements are fully connected and causally heterogeneous (self-connections not drawn). All elements are spontaneously active (1) with heterogeneous probabilities: p(A/D/G) = 0.45; p(B/E/H) = 0.5; p(C/F/I) = 0.55. The elements are structured into three groups {ABC, DEF, GHI} due to different intra- group and intergroup mechanisms: within each group, if the sum of intragroup connections Σ(intra) = 0 (for two time steps), all elements stay 0 (for the next two time steps). However, if the sum of intergroup connections Σ(inter) = 6 from one or both of the other two groups over two time steps (burst of synchronous activity), p(1) is raised by 0.5 for the next two time steps (see Fig. S2 for macro and micro TPMs of a spatial system with equivalent rules). At the macro-level S_M (Fig. 6*B*), the three groups of neurons become macro elements, and two micro time steps (t_x) are grouped into one macro time step (T_x). In neural terms, these macro elements could represent "minicolumns" having three states: "inhibited" (all minicolumn neurons silent at T_x), "receptive" (some firing at T_x), or "bursting" (all firing at T_x). Macro causal interactions can be summarized as follows: if a macro element is inhibited, only receiving a burst can move it to the receptive or (more unlikely) the bursting state; otherwise, it stays inhibited. As in previous examples, the coarse-grained S_M has higher $EI(SM) = 3.51$ bits and $Eff(SM) = 0.74$ than S_m [$EI(Sm) = 0.59$ bits; $Eff(S_m) = 0.033$]. In this case, spatiotemporal causal emergence [$CE(S) = 2.92$ bits] is due to an increase in determinism that far outweighs a slight increase in degeneracy and the decrease in size.

Discussion

This paper provides a principled way of assessing at which spatio-temporal grain size the causal interactions within a system reach a maximum. Causal interactions

are evaluated by effective information (*EI*), a measure that is sensitive both to the effectiveness of the system's mechanisms and to the size of its state space. Examples with simulated systems demonstrate that, after coarse graining the micro mechanisms in both space and time, *EI* can be higher at a macro level than at a micro level. In these cases, the macro mechanisms, rather than the micro ones, can be said to be doing the causal work within a system.

Effective Information, Effectiveness, and Emergence. As shown here, *EI* corresponds to the "effectiveness" of a system's mechanisms multiplied by repertoire size, expressed in bits. Effectiveness *Eff*(*S*) is the average of the effect coefficients over all system states. The effect coefficient measures to what extent the current system state is necessary to specify the system's future state. This, in turn, is a function of determinism minus degeneracy. On the cause side, the equivalent to the effect coefficient is the cause coefficient, which measures to what extent the current state is sufficient to specify the system's past state. For a particular current state, cause and effect coefficients may differ: for example, a state may have many causes but only one effect. However, the average of the effect coefficients over system states, i.e., effectiveness, corresponds to the average of the cause coefficients (weighted by the probability of the effects). In other words, within a time- invariant system the average selectivity of the causes corresponds to the average selectivity of the effects. Note that, in principle, other measures of causation that, like *EI,* reflect causal

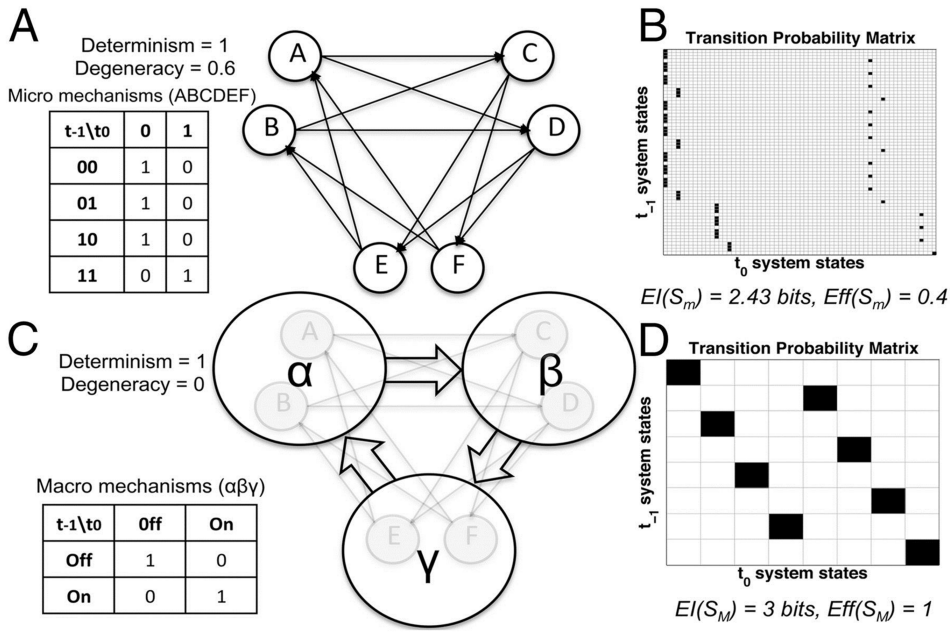


Fig. 4. Spatial causal emergence (counteracting degeneracy). (*A*) A degenerate S_m with deterministic AND gates. (*B*) The cycle of AND gates is mapped onto a cycle of COPY gates at the macro level. (*C*) The deterministic but degenerate micro TPM. (*D*) The deterministic macro TPM with zero degeneracy. By eliminating degeneracy and achieving perfect effectiveness, the macro beats the micro (*CE* = 0.57 bits).

structure (selectivity, determinism, degeneracy) and system size, should demonstrate causal emergence as well.

The main result obtained in the simulations is that coarse graining, both in space and in time, can yield a higher value of *EI*. This happens even though the micro has, by definition, a larger state space than the macro—an advantage with respect to *EI*. Given this inherent advantage of the micro, it is understandable why the default scientific strategy for analyzing systems has been one of reduction (*Causal Reduction*). However, the examples presented above show that the inherent loss in *EI* due to the macro's smaller repertoire size can be offset if the macro achieves a greater gain in effectiveness. In turn, greater effectiveness stems from macro mechanisms constructed from their constituting micro mechanisms in such a way that, at the macro level, determinism is increased and/or degeneracy is decreased. Genuine causal emergence can then be said to occur whenever there is a gain in *EI* ($CE > 0$) at the optimal macro level. If instead there is a loss in *EI* ($CE < 0$), causal reduction is appropriate, and the micro level is the optimal level of causal analysis. The causal approach pursued here suggests that qualitative or noncausal accounts of

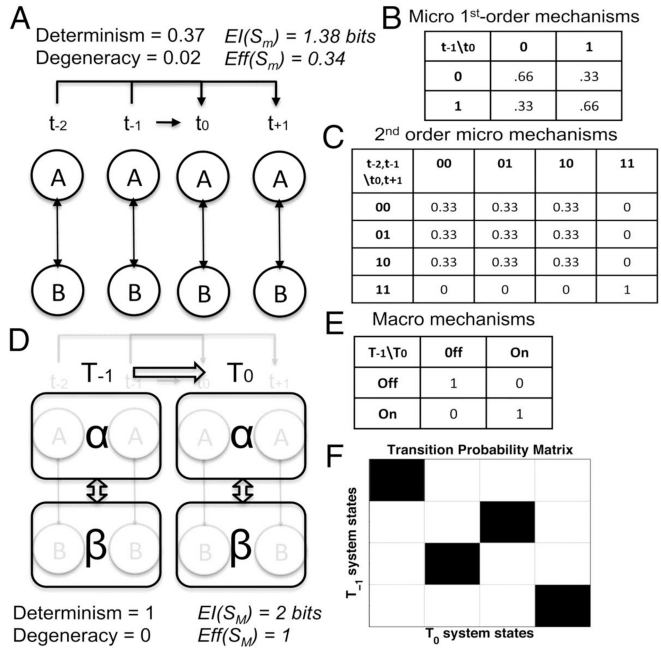


Fig. 5. Temporal causal emergence. (*A*) S_m is composed of second-order Markov mechanisms A and B: at t_0, each mechanism responds based on the inputs at t_{-2} and t_{-1}, and outputs over t_0 and t_{+1}. (*B*) Causal analysis over one micro time step gives an incomplete view of the system. (*C*) A causal analysis over two micro time steps reveals the second-order Markov mechanisms. (*D*) The optimal macro system S_M groups two micro time steps into one macro time step for macro elements $\{\alpha,\beta\}$. (*E*) Each coarse grained macro mechanism effectively corresponds to a deterministic COPY gate. (*F*) The macro one-time step TPM S_M has $Eff(S_M) = 1$, and the micro two-time step TPM has $Eff(S_m) = 0.34$; $CE = 0.62$ bits.

emergence may have been hindered by not being able to characterize how and why a macro level can actually have greater causal effectiveness than a micro level (22, 23).

Micro Macro Mappings and Repertoires of Alternatives. The present approach makes it possible to compare causation at the micro and macro levels in a fair manner. First, the simulated examples are such that the macro supervenes strictly upon the micro: once the micro is defined, all macro levels are fixed. Specifically, no extra causal ingredients are added at the macro level, such as rules that apply to the macro only (24). Furthermore, the mapping of micro into macro elements is such that the identity of micro elements is lost; otherwise, the macro level would have access to micro-level information that could offset its reduced repertoire size. Finally, when causation is evaluated a uniform distribution of alternatives is imposed independently at the micro and macro levels. For this uniform distribution of perturbations to be imposed at the macro level, the probability of the underlying micro perturbations must be modified by averaging the micro states that map into the same macro state. The modified distribution of micro perturbations yielding a uniform distribution of macro perturbations makes *EI* sensitive to the causal structure at each level, ultimately allowing the supervening macro *EI* to exceed the micro *EI*.

Emergence as an Intrinsic Property of a System. *EI* is a causal measure, because it requires perturbing the system in all possible ways and evaluating the resulting effects on the system. It is also an informational measure, because its value depends on the size of the repertoire of alternatives. Indeed, in the present approach, causation and information are necessarily linked (25), hence the term "effective information." Finally, measuring *EI* reveals an "intrinsic" property of the system, namely the average effectiveness/selectivity of all possible system states with respect to the system itself. Effectiveness/selectivity can be assessed at multiple spatiotemporal grains, and the particular spatiotemporal grain at which *EI* reaches a maximum is again an intrinsic property of the system. This in no way precludes an observer from profitably investigating the system's properties at other macro levels, at the micro level, or at multiple levels at once (e.g., neuroscientists studying the brain at the level of ion channels, individual neurons, local field potentials, or functional magnetic resonance signals). However, causal emergence implies that the macro level with highest *EI* is the one that is optimal to characterize, predict, and retrodict the behavior of the system—the one that "carves nature at its joints" (26).

The search for the macro level at which *EI* is maximal has a parallel in information theory: channel capacity is an intrinsic property defined as the maximal amount of information that can be transmitted along the channel at a certain rate, found by searching over all possible input distributions (27). Finding the optimal level of coarse graining for causal emergence is based on a similar search, with several differences. First, *EI* is evaluated using perturbations over the system itself, rather than across a channel (the system is its own input and output). Second, the probability distributions over micro states that can be considered must conform to a proper mapping of micro into macro elements (or time intervals). Additional

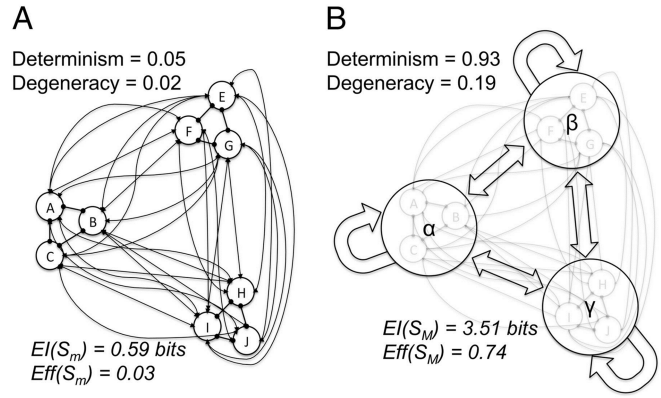


Fig. 6. Spatiotemporal causal emergence. (*A*) A "neuronal" system merging the temporal characteristics of the system in Fig. 5 with a differentiated spatial structure (Fig. S2). Regular and rounded arrows indicate intergroup and intragroup connections, respectively. (*B*) Each macro element receives inputs from itself and the other macro element. The macro level beats the micro level, leading to spatiotemporal emergence [$CE(S) = 2.92$ bits].

connections of causal emergence to established measures, such as reversibility and lumping in Markov processes (28), or epsilon machines (29), are a potential subject for future work.

Causal Exclusion and Its Implications. Causal analysis as presented here endorses both supervenience (no extra causal ingredients at the macro level) and causal exclusion [for a given system at a given time, causation occurs at one level only, otherwise causes would be double counted (4)]. However, causal analysis also demonstrates that *EI* can actually be maximal at a macro level, depending on the system's architecture. In such cases, causal exclusion turns the reductionist assumption on its head, because to avoid double-counting causes, optimal macro causation must exclude micro causation. In other words, macro mechanisms can always be decomposed to their constituting micro mechanisms (supervenience); however, if there is emergence, macro causation does not reduce to micro causation, in which case the macro wins causally against the micro and takes its place (supersedence). The notion of irreducibility among levels (does the macro beat the micro?) is complemented by the notion of irreducibility among subsets of elements within a level [is the whole more than its parts (15, 25)?]. From the perspective of a system, emergence ($CE > 0$) implies causal "self-definition" at the optimal macro level—the one at which its causal interactions "come into focus" (30) and "the action happens."

Applicability to Real Systems. Measuring *EI* exhaustively, across all micro/macro levels, is not feasible for complex physical or biological systems (*Applicability—Network Motifs as Indicators of Emergence*). However, some useful guidelines can be derived from the above analysis: (*i*) if $Eff(S_m) \geqslant Eff(S_M)$, then causal emergence is impossible and causal reduction holds; (*ii*) if $EI(S_m) > \log_2(n_M)$, where n_M is the state

repertoire size of S_M, causal reduction holds; (*iii*) if for some coarse graining, *Eff* increases drastically, causal emergence is to be suspected (as $\Delta I_{Eff} >> -\Delta I_{Size}$). Therefore, systems that already are close to maximal effectiveness at the micro level (Fig. S1) indicate causal reduction. By contrast, heavily interconnected groups of elements with spontaneous activity and the ability to distinguish between intragroup and intergroup connections, such as the simplified neural system of Fig. 6, are more suitable for emergence.

In real neural systems, one could compare the respective effective information at the micro scale of single neurons over millisecond intervals, the meso scale of neuronal groups over hundreds of milliseconds, and the macro scale of brain regions over several seconds (using tools such as optogenetics and calcium imaging). In this way, classic notions, such that cortical minicolumns may constitute the fundamental units of brain function (31), or that the cortex works by population coding in space (32) or rate coding in time (33) in the face of high intertrial variability (34), could then be tested rigorously using a measure of effectiveness. Examining small motifs that are overrepresented in complex networks [such as brains (35)] could determine whether the network as a whole is biased toward emergence or reduction. Heuristic assessments of the likelihood of emergence could also rely on the analysis of wiring diagrams, which can offer an estimate of degeneracy, combined with knowledge of the amount of intrinsic noise in a system, which can provide an estimate of determinism.

Conclusions

The approach to emergence investigated here provides theoretical support for the intuitive idea that, to find out how a system works, one should find the "differences that make [most of] a difference" to the system itself (25) (cf. ref. 36). It also suggests that complex, multilevel systems such as brains are likely to "work" at a macro level because, in biological systems, selectional processes must deal with unpredictability and lead to degeneracy (18). This may also apply to some engineered systems designed to compensate for noise and degeneracy. More broadly, this view of causal emergence suggests that the hierarchy of the sciences, from microphysics to macroeconomics, may not just be a matter of convenience but a genuine reflection of causal gains at the relevant levels of organization.

ACKNOWLEDGMENTS. We thank M. Boly, C. Cirelli, A. Hashmi, C. Koch, L. Morton, A. Nere, M. Oizumi, and L. Shapiro for helpful discussions, and P. Rana for assisting with the Python software. This work has been supported by Defense Advanced Research Planning Agency Grant HR 0011-10-C-0052 and The Paul G. Allen Family Foundation.

1. Sporns O, Tononi G, Kotter R (2005) The human connectome: A structural description of the human brain. *PLoS Comput Biol* 1(4):e42.
2. Markram H (2006) The blue brain project. *Nat Rev Neurosci* 7(2):153–160.
3. Davidson D (1980) Mental events. *Readings in Philosophy of Psychology,* ed Block N (Harvard Univ Press, Cambridge, MA), Vol 1, pp 107–119.

4. Kim J (1993) *Supervenience and Mind: Selected Philosophical Essays* (Cambridge Univ Press, Cambridge, UK).
5. Kim J (2000) *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation* (MIT Press, Cambridge, MA).
6. Bontly T (2002) The supervenience argument generalizes. *Philos Stud* 109:75–96.
7. Seth A (2008) Measuring emergence via nonlinear Granger causality. *ALIFE* 2008:545–552.
8. Hölldobler B, Wilson E (2009) *The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies* (W. W. Norton, New York).
9. Sperry R (1983) *Science and Moral Priority: Merging Mind, Brain, and Human Values* (Columbia Univ Press, New York).
10. Sawyer R (2005) *Social Emergence: Societies as Complex Systems* (Cambridge Univ Press, Cambridge, UK).
11. Broad C (1925) *The Mind and Its Place in Nature* (Routledge & Kegan Paul, London).
12. Bar-Yam Y (2004) A mathematical theory of strong emergence using multiscale variety. *Complexity* 9:15–24.
13. Tononi G, Sporns O (2003) Measuring information integration. *BMC Neurosci* 4:31.
14. Pearl J (2000) *Causality: Models, Reasoning and Inference* (Cambridge Univ Press, Cambridge, UK).
15. Albantakis L, Hoel EP, Koch C, Tononi G (2013) Intrinsic Causation and Consciousness. *Association for the scientific study of consciousness conference (ASSC17).* Available at www.theassc.org/files/assc/docs/ASSC17-PB-070113-online-version-with-Addendum.pdf. Accessed November 2, 2013.
16. Kullback S (1997) *Information Theory and Statistics* (Dover Publications, New York).
17. Edelman GM (1987) *Neural Darwinism: The Theory of Neuronal Group Selection* (Basic Books, New York).
18. Tononi G, Sporns O, Edelman GM (1999) Measures of degeneracy and redundancy in biological networks. *Proc Natl Acad Sci USA* 96(6):3257–3262.
19. Stalnaker R (1996) Varieties of supervenience. *Philos Perspect* 10:221–241.
20. Fodor J (1974) Special sciences (or: The disunity of science as a working hypothesis). *Synthese* 28:97–115.
21. Jahr CE, Stevens CF (1990) Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics. *J Neurosci* 10 (9):3178–3182.
22. Bedau M (1997) Weak emergence. *Nous* 31:375–399.
23. Chalmers D (2006) Strong and weak emergence. *The Reemergence of Emergence,* eds Clayton P, Davies P (Oxford Univ Press, Oxford), pp 244–256.
24. Butterfield J (2012) Laws, causation and dynamics at different levels. *Interface Focus* 2(1):101–114.
25. Tononi G (2012) Integrated information theory of consciousness: An updated account. *Arch Ital Biol* 150(2-3):56–90.

26. Hamilton E, Cairns H (1961) *The Collected Dialogues of Plato: Including the Letters* (Pantheon Books, New York).
27. Shannon CE (1997) The mathematical theory of communication. 1963. *MD Comput* 14(4):306–317.
28. Kemeny J, Snell J (1976) *Finite Markov Chains* (Springer, New York).
29. Shalizi C, Crutchfield J (2001) Computational mechanics: Pattern and prediction, structure and simplicity. *J Stat Phys* 104:817–879.
30. Alexander S (1920) *Space, Time, and Deity: The Gifford Lectures at Glasgow, 1916–1918* (Macmillan, London).
31. Buxhoeveden DP, Casanova MF (2002) The minicolumn hypothesis in neuroscience. *Brain* 125(Pt 5):935–951.
32. Georgopoulos AP, Schwartz AB, Kettner RE (1986) Neuronal population coding of movement direction. *Science* 233(4771):1416–1419.
33. London M, Roth A, Beeren L, Hausser M, Latham PE (2010) Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature* 466(7302):123–127.
34. Knoblauch A, Palm G (2005) What is signal and what is noise in the brain? *Biosystems* 79(1-3):83–90.
35. Sporns O (2010) *Networks of the Brain* (MIT Press, Cambridge, MA).
36. Fitelson B, Hitchcock C (2010) *Probabilistic Measures of Causal Strength* (Oxford Univ Press, Oxford, UK).

Supporting Information

Hoel etal. 10.1073/pnas.1314922110

Effect Coefficient and Effectiveness *(Eff)* Expressed as Determinism and Degeneracy

The state-dependent effect coefficient $(s_0) = \dfrac{\text{effect information}(s_0)}{\log_2(n)}$ can be described as a function of two terms, the determinism and degeneracy coefficient. To derive these two terms, the effect information (s_0), the distance between the effect repertoire $(S_F|s_0)$ and the unconstrained repertoire of effects U^E, is split into the distance between $(S_F|s_0)$ and the uniform distribution U with $p(s_U) = 1/n$, and a residual term:

$$\text{Effect Information}(s_0) = D_{\text{KL}}((S_F|s_0),\ U^E) = \sum_{S_F \in U^E} p(S_F|s_0)\log_2\left(\frac{p(S_F|s_0)}{p(s_F)}\right) \tag{S1}$$

$$= \sum_{s_F \in U^E} p(s_F|s_0)\log_2\left(\frac{p(s_F|s_0)}{p(s_U)} + \frac{p(S_U)}{p(S_F)}\right) \tag{S2}$$

$$= \sum_{s_F \in U^E} p(s_F|s_0)\left(\log_2\left(\frac{p(s_F|s_0)}{p(s_U)}\right) - \log_2\left(\frac{p(S_F)}{p(S_U)}\right)\right) \tag{S3}$$

$$= \sum_{s_F \in U^E} p(s_F|s_0)\log_2\left(\frac{p(s_F|s_0)}{p(s_U)}\right) - \sum_{s_F \in U^E} p(s_F|s_0)\log_2\left(\frac{p(s_F)}{p(s_U)}\right) \tag{S4}$$

$$(\text{using } p(s_U) = 1/n) = \sum_{s_F \in U^E} p(s_F|s_0)\log_2(n \cdot p(s_F|s_0)) - \sum_{s_F \in U^E} p(s_F|s_0)\log_2(n \cdot p(s_F)) \tag{S5}$$

$$= D_{\text{KL}}((S_F|s_0),\ U) - \sum_{s_F \in U^E} p(s_F|s_0)\log_2(n \cdot p(s_F)), \tag{S6}$$

where s_F denotes a state of the system S_F at t_{+1} with probability $p(s_F)$ according to the unconstrained distribution of effects $U^E \cdot s_0$ is the present system state. The determinism coefficient is then the left term in lines **S5** and **S6** divided by log2(n):

$$\text{Degeneracy coefficient}(S_0) = \frac{\sum_{s_F \in U^E} p(S_F|s_0)\log_2(n \cdot p(S_F|s_0))}{\log_2(n)} \tag{S7}$$

the degeneracy coefficient the right term:

$$\text{Degeneracy coefficient}(s_0) = \frac{\sum_{s_F \in U^E} p(s_F|s_0)\log_2(n \cdot p(s_F|s_0))}{\log_2(n)} \tag{S8}$$

as defined in the main article.

The effectiveness (*Eff*) of a system assesses the causal relations in a system in a state-independent manner, irrespective of the size of the system's state space:

$$\begin{aligned} Eff(S) &= \frac{EI(S)}{\log_2(n)} = \frac{\langle \text{Effect Information}(s_0)\rangle}{\log_2(n)} \\ &= \frac{\sum_{s_0 \in U^C} p(s_0) D_{\mathrm{KL}}((S_F|s_0),\ U^E)}{\log_2(n)}, \end{aligned} \tag{S9}$$

where the effective information $EI(S)$ is the average effect information of all system states s_0, distributed according to U^C, the unconstrained repertoire of causes, which is identical to the uniform distribution U; thus, here $p(s_0) = 1/n$. $EI(S)$ can then be divided in the same way as the state-dependent effect information:

$$EI(S) = \langle \text{Effect Information}(s_0)\rangle, \tag{S10}$$

$$= \left\langle D_{\mathrm{KL}}((S_F|s_0),\ U) - \sum_{s_F \in U^E} p(S_F|s_0)\log_2\left(\frac{p(s_F)}{p(s_U)}\right)\right\rangle, \tag{S11}$$

$$= \langle D_{\mathrm{KL}}((S_F|s_0),\ U)\rangle - \left\langle \sum_{s_F \in U^E} p(S_F|s_0)\log_2\left(\frac{p(s_F)}{p(s_U)}\right)\right\rangle, \tag{S12}$$

$$= \langle D_{\mathrm{KL}}((S_F|s_0),\ U)\rangle - \sum_{s_0 \in U^C} p(s_0) \sum_{s_F \in U^E} p(s_F|s_0)\log_2\left(\frac{p(s_F)}{p(s_U)}\right), \tag{S13}$$

$$= \langle D_{\mathrm{KL}}((S_F|s_0),\ U)\rangle - \sum_{s_F \in U^E} p(s_F)\log_2\left(\frac{p(s_F)}{p(s_U)}\right), \tag{S14}$$

$$= \langle D_{\mathrm{KL}}((S_F|s_0),\ U)\rangle - D_{\mathrm{KL}}(U^E,\ U). \tag{S15}$$

The last equality is due to the fact that $p(s_F)$ is the probability of state s_F to occur at t_{+1} following U^E, the unconstrained distribution of effects (future states) obtained by setting the system S at t_0 into all possible states s_0 with equal probability $p(s_0) = 1/n$.

Both, indeterminism and degeneracy at the micro level may be indicative of causal emergence (*Discussion,* main text). Note that, in previous work, it was suggested that a convergence of two causes onto the same effect—an instance of

degeneracy—may actually disqualify the micro level from causation (1, 2) (although see ref. 3).

1. Yablo S (1992) Mental causation. *Philos Rev* 101:245–280.
2. List C, Menzies P (2009) Non-reductive physicalism and the limits of the exclusion principle. *J Philos* CVI(9):475–502.
3. Shapiro L, Sober E (2012) Against proportionality. *Analysis* 72:89–93.

Effective Information *EI*(*S*) Expressed in Terms of Cause and Effect Information and Mutual Information *MI*

The effective information of a system, $EI(S)$, can be obtained as the expected value of the cause or effect information. Moreover, $EI(S)$ is identical to the mutual information $MI(U^C; U^E)$: the MI between the system S set to all possible counterfactuals (system states) with equal probability (unconstrained repertoire of causes, U^C) and the resulting distribution of system states at the next time step (unconstrained repertoire of effects, U^E). Note that EI was originally introduced as a measure of causal influence of one subset of a system over another (1), whereas here it captures the overall effectiveness of system S onto itself (see refs. 2 and 3 for related measures).

In the following derivation, we start from the definition of $EI(S)$ as the average effect information of all system states s_0 as counterfactual causes [distributed according to U^C with equal probability $p(s_0) = 1/n$ for all system states]:

$$\mathrm{EI}(S) = \langle \text{Effect Information}(s_0) \rangle = \sum_{s_0 \in U^C} p(s_0) D_{\mathrm{KL}}((S_F|s_0),\ U^E) = \tag{S1}$$

$$(\text{using } p(s_0) = 1/n\ \forall\ s_0) = \frac{1}{n} \sum_{s_0 \in U^C} D_{\mathrm{KL}}((S_F|s_0),\ U^E). \tag{S2}$$

Using Bayes' rule and time invariance, we then show that the average effect information is indeed equivalent to the mutual information $MI(U^C; U^E)$ and to the expected value of the cause information, which is the average cause information of each accessible state at t_0, weighted by $p(s_0)$ according to U^E:

$$\begin{aligned} EI(S) &= \langle \text{Effect Information}(s_0) \rangle = MI(U^C; U^E) \\ &= \langle \text{Cause Information}(s_0) \rangle. \end{aligned} \tag{S3}$$

In detail:

$$EI(S) = \langle \text{Effect Information}(s_0) \rangle = \sum_{s_0 \in U^C} p(s_0) D_{\mathrm{KL}}((S_F|s_0),\ U^E) = \tag{S4}$$

$$= \sum_{s_0 \in U^C} p(s_0) \sum_{s_F \in U^E} p(s_F|s_0) \log_2 \left(\frac{p(s_F|s_0)}{p(s_F)} \right) = \tag{S5}$$

$$= \sum_{s_0 \in U^C} \sum_{s_F \in S^F} p(s_0) p(s_F|s_0) \log_2 \left(\frac{p(s_F|s_0)}{p(s_F)} \right) = \tag{S6}$$

$$(\text{Bayes}' \text{ rule}) = \sum_{s_0 \in U^C} \sum_{s_F \in U^E} p(s_0|s_F) \log_2 \left(\frac{p(s_0, s_F)}{p(s_0) p(s_F)} \right) = \tag{S7}$$

$$= MI(U^C;\ U^E) = \tag{S8}$$

$$(\text{time invariance}) = \sum_{s_P \in U^C} \sum_{s_0 \in U^E} p(s_P|s_0) \log_2 \left(\frac{p(s_P, s_0)}{p(s_P) p(s_0)} \right) = \tag{S9}$$

$$(\text{Bayes}' \text{ rule}) = \sum_{s_P \in U^C} \sum_{s_0 \in U^E} p(s_0) p(s_P|s_0) \log_2 \left(\frac{p(s_P, s_0)}{p(s_P)} \right) = \tag{S10}$$

$$= \sum_{s_0 \in U^E} p(s_0) \sum_{s_P \in U^C} p(s_P|s_0) \log_2 \left(\frac{p(s_P|s_0)}{p(s_P)} \right) = \tag{S11}$$

$$= \sum_{s_0 \in U^E} p(s_0) D_{\mathrm{KL}}((s_P|s_0),\ U^C) = \langle \text{Cause Information}(s_0) \rangle. \tag{S12}$$

MI is originally a statistical measure of how much information is shared between a source and a target (4). In the present context, *MI* is applied between two time steps of a system that is first perturbed into all counterfactuals (alternative states) with equal probability and then observed at the next time step. Because of the system perturbations, *MI* here is a causal measure. In other words, $EI(S)$ is the *MI* between the set of all possible causes U^C and the set of all their effects U^E. Usually, however, *MI* is calculated for observed distributions of system states and thus not a causal measure, but a statistical measure of correlation.

1. Tononi G, Sporns O (2003) Measuring information integration. *BMC Neurosci* 4:31.
2. Ay N, Polani D (2008) Information flows in causal networks. *Adv Complex Syst* 11(1): 17–41.
3. Korb KB, Nyberg EP, Hope L (2011) *Causality in the Sciences,* eds Illari P, Russo F, Williamson J (Oxford Univ Press, Oxford), pp 628–652.
4. Cover TM, Thomas JA (2006) *Elements of Information Theory* (Wiley-Interscience, Hoboken, NJ).

Bounds of Cause and Effect Coefficients and Effectiveness *Eff*(*S*)

In the following, we will show that the cause and effect coefficients, as well as the effectiveness $Eff(S)$, are bounded between 0 and 1 ($\in [0 \ldots 1]$):

$$\begin{aligned}\text{Cause coefficient}(s_0) &= \frac{\text{Cause information}(s_0)}{\log_2(n)} \\ &= \frac{D_{\mathrm{KL}}((S_P \mid s_0),\ U^C)}{\log_2(n)},\end{aligned} \tag{S1}$$

$$\begin{aligned}\text{Effect coefficient}(s_0) &= \frac{\text{Effect information}(s_0)}{\log_2(n)} \\ &= \frac{D_{KL}((S_F \mid s_0),\ U^E)}{\log_2(n)},\end{aligned} \tag{S2}$$

$$\begin{aligned}Eff(S) = \frac{EI(S)}{\log_2(n)} &= \frac{\frac{1}{n}\sum_{s_0 \in U^C} D_{KL}((S_F \mid s_0),\ U^E)}{\log_2(n)} \\ &= \langle \text{Effect cofficient}(s_0) \rangle.\end{aligned} \tag{S3}$$

The lower bound (0) is given by the fact that the Kullback–Leibler divergence (D_{KL}) is always nonnegative (Gibbs' inequality). Because the cause and effect information are expressed in terms of D_{KL} and the state-independent effective information $EI(S)$ is just an average of the state-dependent values, neither of the three coefficients can be negative. It thus remains to show that cause and effect coefficients cannot exceed 1.

The cause information (s_0) is the D_{KL} between the cause repertoire ($S_p|s_0$) and U^C, the unconstrained cause repertoire, which is identical to the uniform distribution with $p(s_p) = 1/n\ \forall s_P$. It follows that

$$\begin{aligned}\text{Causeinformation}(s_0) &= D_{\mathrm{KL}}((S_P \mid s_0),\ U^C) \\ &= \sum_{s_P \in U^C} p(s_P \mid s_0) \log_2\left(\frac{p(s_P \mid s_0)}{p(s_P)}\right) =\end{aligned} \tag{S4}$$

$$= \sum_{s_P \in U^C} p(s_P \mid s_0) \log_2(n \cdot p(s_P \mid s_0)) = \tag{S5}$$

$$(\text{since } p(s_P \mid s_0) \leqslant 1) \leqslant \sum_{s_P \in U^C} p(s_P \mid s_0) \log_2(n) = \log_2(n), \tag{S6}$$

and thus

$$\text{Cause coefficient}(s_0) \leqslant 1. \tag{S7}$$

The effect information (s_0) is the D_{KL} between the effect repertoire ($S_F|s_0$) and U^E, the unconstrained effect repertoire. U^E is in general not identical to the uniform distribution. However,

$$p(s_F) = \sum_{s_0 \in U^C} p(s_F \mid s_0) \cdot p(s_0), \tag{S8}$$

where $p(s_0) = 1/n \; \forall s_0$ and thus:

$$p(s_F|s_0) \leqslant n \cdot p(s_F), \; \forall \; s_F. \tag{S9}$$

Using Eq. **S9**, if follows that:

$$\begin{aligned} \text{Effect information}(s_0) &= D_{\mathrm{KL}}((S_F|s_0), \; U^E) \\ &= \sum_{s_F \in U^E} p(s_F|s_0) \log_2\left(\frac{p(s_F|s_0)}{p(s_F)}\right) = \end{aligned} \tag{S10}$$

$$(\text{using Eq. } \mathbf{S9} \leq \sum_{s_F \in U^E} p\left(s_F|s_0\right) \log_2\left(\frac{n \cdot p(s_F)}{p(s_F)}\right) = \sum_{s_F \in U^E} p\left(s_F|s_0\right) \log_2(n) \tag{S11}$$

$$= \log_2(n), \tag{S12}$$

and thus

$$\text{Effect coefficient}(s_0) \leqslant 1. \tag{S13}$$

Finally, because the effect coefficient $(s_0) \in [0 \ldots 1] \; \forall s_0$, also its average over all system states, the state-independent effectiveness $Eff\,(S) \in [0 \ldots 1]$.

Causal Reduction

To complement the examples of causal emergence in the main text, we here provide an example in which causal reduction is called for. In Fig. S1, a macro mechanism works as an XOR logic gate (as an isolated part of a larger circuit board) with inputs X, Y, and output Z (Fig. S1*A*). At the macro level, the system (XOR,X,Y,Z) generates 2 bits of *EI* over one macro time step T_x (the XOR operates after a "decision" period where it processes the input) and $Eff(S_M) = 0.5$. The macro XOR gate is actually composed of (supervenes upon) nine deterministic micro logic gates (COPY, NOT, AND, OR). In this case, however, causal interactions are stronger at the micro level and over a single micro time step $t_x[EI(SM) = 7.43$ bits and $Eff(s_M) = 0.83$]. Thus, $CE = -5.43$ bits, corresponding to negative causal emergence, i.e., reduction. Note that in this case the micro circuit is deterministic and minimally degenerate (0.17), so the macro cannot offset the loss of effective information due to its reduced size by a gain in determinism or a reduction in degeneracy.

To demonstrate this case of causal reduction, we have assumed that a deterministic micro circuit underlies the above macro circuit. In general, however, real digital circuits are often built from many stochastic analog micro elements in a highly degenerate manner, to compensate for noise at the lower level and to create deterministic macro elements. In this way, digital circuits and other engineered systems follow similar design principles as the more physiological examples presented in the main text. Consequently, there is the potential for either causal

emergence or reduction in digital circuits, depending on the underlying micro level, just as in physiological systems.

More generally, the notion of causal reduction ($CE < 0$) stands in contrast to previous accounts of reduction that focused on the relationship between scientific theories and whether or not they are reducible to one another (1). In the present account based on causal analysis, the focus is instead on the relationship between micro and macro levels of mechanisms. This account reveals why there is a bias in favor of reductionism in mechanistic scientific explanations. The bias is understandable given that, everything else being equal, the micro would always beat the macro: being more detailed by definition, the micro has an inherent advantage in how informative its causal mechanisms are. This inherent advantage is captured quantitatively in causal analysis because the micro can benefit from both ΔI_{Eff} and ΔI_{Size}, whereas the macro can only gain from ΔI_{Eff}.

1. Nagel E (1961) *The structure of science: problems in the logic of scientific explanation* (Harcourt, Brace & World, New York).

Causal Emergence in a System with Causally Heterogeneous Elements

Although the examples in the main text (with the exception of Fig. 6) all have macro elements with underlying unconnected and causally equivalent micro elements, this is not a necessity for causal emergence. In Fig. S2*A*, the six micro elements are fully interconnected and causally heterogeneous. The elements are structured into two groups {ABC, DEF} due to different intra-group and intergroup mechanisms: within each group, if the sum of intragroup connections Σ(intra) = 0, all elements stay 0 (inactive) the next time step. However, if the sum of intergroup connections Σ(inter) = 3 (synchronous activity from the other group), all elements turn 1, unless they are all 0, in which case they become spontaneously active (1) with probabilities: p(A/D) = 0.45; p(B/ E) = 0.5; p(C/F) = 0.55. Because the micro transition probability matrix (TPM) is noisy, $EI(S_m) = 1.13$ bits and $Eff(S_m) = 0.19$ (Fig. S2*B*). The optimal macro grouping S_M (Fig. S2*C*) has a more deterministic TPM (Fig. S2*D*), $EI(S_M) = 1.84$ bits and $Eff(S_M) = 0.58$. Thus, the macro supersedes the micro [$CE(S) = 0.72$ bits] despite its reduced repertoire size, because it counteracts noise by responding almost deterministically to synchronous activity over intergroup connections.

The neural-like system of Fig. 6 in the main text has equivalent spatial properties to the example system of Fig. S2 (fully connected, causally heterogeneous elements, sensitive to differences in intraconnections and interconnections). In addition, it has the same temporal properties as the system shown in Fig. 5 (main text), with second-order Markov mechanisms at the micro level. The system's states space at the micro level thus contains 2^{18} states, which prohibited an exhaustive search for the optimal macro level. Nevertheless, the spatiotemporally emergent macro grouping shown in Fig. *6B* (main text) is assumed to be the optimal macro grouping based on the results obtained from the examples of Fig. S2 and Fig. 5 (main text).

Applicability—Network Motifs as Indicators of Emergence

Measuring *EI* exhaustively, across all micro/macro levels, is not feasible for large systems. This is because, assuming N binary elements, $B_N - 1$ (Nth Bell number) possible groupings of those micro elements into macro elements exist, each of which entails $\prod_{j=1}^{k} (B_{m(j)+1} - 1)$ possible groupings of micro into macro states, where k is the number of macro elements with $m(j)$ micro elements each. The number of *EI* computations to determine the spatiotemporal grain with maximal *EI* thus increases dramatically with N ($N = 1$, 1; $N = 2$, 5; $N = 3$, 27; $N = 4$, 180 computations, etc.) if calculated exhaustively.

In large, complex networks where an exhaustive causal analysis is unfeasible, overrepresented network motifs could already indicate whether the network as a whole is biased toward emergence or reduction. For example, the two most common network motifs shared by the gene networks in *Escherichia coli* and the brain of *Caenorhabditis elegans* are the feedforward loop and the bifan (1). Both these network motifs mimic in their connectivity precisely the micro element groups that made up the optimal (winning) macro elements in our chosen examples. In Fig. 2 (main text), the first spatial example, the macro elements are bifans, whereas in Fig. 6 (main text), the first temporal example, the macro elements are feedforward loops. These are perhaps the simplest possible functionally relevant macro elements. Both the bifan and the feedforward loop show causal convergence (degeneracy) in either space or time. A greater than random prevalence of these or similar network motifs, paired with some amount of intrinsic noise in the system, may indicate that the system operates at a macro level.

1. Milo R, et al. (2002) Network motifs: Simple building blocks of complex networks. *Science* 298(5594):824–827.

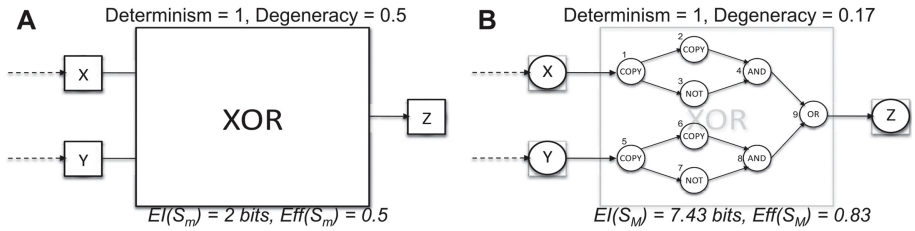


Fig. S1. Causal reduction. (*A*) A part of a larger circuit is presented, which performs a macro XOR logic function over its inputs X, Y, and outputs to Z. (*B*) At the micro level, the XOR consists of nine deterministic logic gates. The system is deterministic at both the micro and the macro level. Moreover, the degeneracy coefficient at the micro level is lower than at the macro level. Therefore, in this case, the micro beats the macro, leading to causal reduction. $CE(S) = -5.43$ bits.

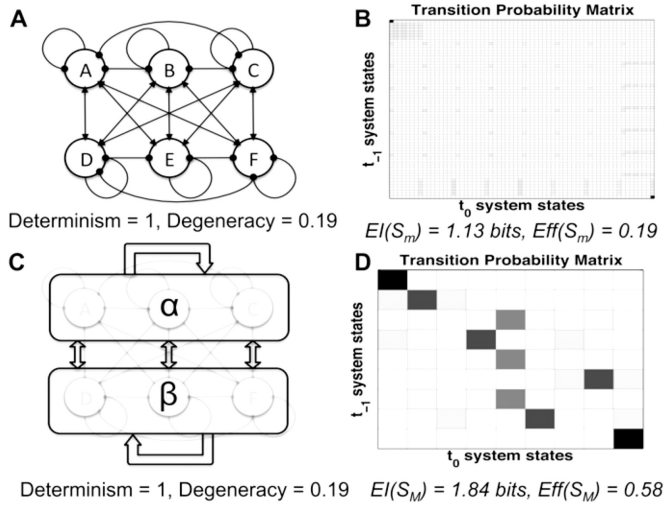
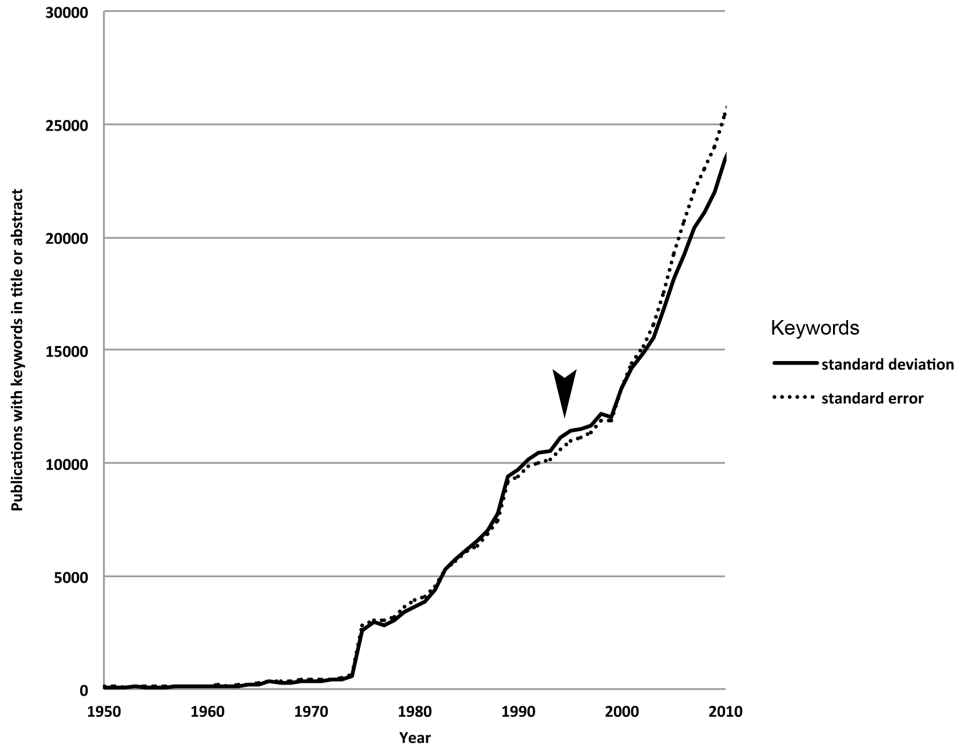


Fig. S2. Causal emergence in a system with differentiated connectivity. (*A*) Micro system S_m with six elements. Regular and rounded arrows indicate intergroup and intragroup connections, respectively. (*B*) Noisy micro-level TPM. (*C*) Macro system S_M. Each macro element receives inputs from itself and the other macro element. (*D*) More deterministic macro-level TPM. $CE(S) = 0.72$ bits.

Case Study 15: Standard Deviation not S.E.M.



Streiner (1996) Maintaining standards: differences between the standard deviation and the standard error, and when to use each.

Perspective on *Maintaining Standards: Differences between the Standard Deviation and Standard Error, and When to Use Each*
Can J. Psychiatry 41:498–502

Dany Spencer Adams
Research Professor
Tufts University

> **'…[U]ncertainty and its components should be expressed in the form of 'standard deviations.'** Recommendation INC-1 (1980) by the Working Group on the Statement of Uncertainties convened by the [International Bureau of Weights and Measures], a recommendation that the [International Committee for Weights and Measures] approved in 1981 and reaffirmed in 1986 via its own Recommendations 1 (CI-1981) and 1 (CI-1986).
>
> **'The SE reflects the variability of the mean values, as if the study were repeated a large number of times. By itself, the SE is not particularly useful…'** (Streiner 1996).

This is not the first paper to introduce or even review these ideas. A PubMed search for articles similar to this one finds 122 papers published since1946, including 16 that were published prior to 1996, and there are surely more that were not recognized by this search technique. This one was chosen for this collection because it is exceptionally clear. The rules it explains were taught to us all but are, nevertheless, often disregarded, probably because dividing by the square root of the sample size makes the uncertainty look smaller; that is not a good reason. The trend graph shows how SE and SD are neck and neck when it comes to use. This article explains why that is unfortunate.

Maintaining Standards: Differences between the Standard Deviation and Standard Error, and When to Use Each

David L Streiner, PhD[1]

Many people confuse the standard deviation (SD) and the standard error of the mean (SE) and are unsure which, if either, to use in presenting data in graphical or tabular form. The SD is an index of the variability of the original data points and should be reported in all studies. The SE reflects the variability of the mean values, as if the study were repeated a large number of times. By itself, the SE is not particularly useful; however, it is used in constructing 95% and 99% confidence intervals (CIs), which indicate a range of values within which the "true" value lies. The CI shows the reader how accurate the estimates of the population values actually are. If graphs are used, error bars equal to plus and minus 2 SEs (which show the 95% CI) should be drawn around mean values. Both statistical significance testing and CIs are useful because they assist the reader in determining the meaning of the findings.

(Can J Psychiatry 1996;41:498-502)

Key Words: *statistics, standard deviation, standard error, confidence intervals, graphing*

Imagine that you've just discovered a new brain protein that causes otherwise rational people to continuously mutter words like "reengineer," "operational visioning," and "mission statements." You suspect that this new chemical, which you call LDE for Language Destroying Enzyme, would be found in higher concentrations in the cerebrospinal fluid (CSF) of administrators than that of other people. Difficult as it is to find volunteers, you eventually get samples from 25 administrators and an equal number of controls and find the results shown in Table I. Because you feel that these data would be more compelling if you showed them visually, you prepare your paper using a bar graph. Just before you mail it off, though, you vaguely remember something about error bars, but can't quite recall what they are; you check with a few of your colleagues. The first one tells you to draw a line above and below the top of each bar so that each part is equal to the standard deviation. The second person disagrees, saying that the lines should reflect the standard errors, while the third person has yet another opinion—the lines should be plus and minus 2 standard errors, that is, 2 standard errors above and 2 below the mean. As you can see in Figure 1, these methods result in very different pictures of

Manuscript received March 1996.
This article is the eleventh in the series on Research Methods in Psychiatry.
For previous articles please see Can J Psychiatry 1990;35:616–20, 1991; 36:357–62, 1993;38:9–13, 1993;38:140–8, 1994;39:135–40, 1994; 39:191–6, 1995;40:60–6, 1995;40:439–44, 1996;41:137–43, and 1996;41:491–7.
[1]Professor, Departments of Clinical Epidemiology and Biostatistics and of Psychiatry, McMaster University, Hamilton, Ontario.
Address reprint requests to: Dr David L Streiner, Department of Clinical Epidemiology and Biostatistics, McMaster University, 1200 Main Street West, Hamilton, ON L8N 3Z5
e-mail: streiner@fhs.csu.mcmaster.ca

what's going on. So, now you have 2 problems: first, what is the difference between the standard error and the standard deviation, and second, which should you draw?

Standard Deviation

The standard deviation, which is abbreviated variously as S.D., SD, or s (just to confuse people), is an index of how closely the individual data points cluster around the mean. If we call each point "X_i," so that "X_1" indicates the first value, "X_2" the second value, and so on, and call the mean "M," then it may seem that an index of the dispersion of the points would be simply $\Sigma(X_i - M)$, which means to sum (that's what the Σ indicates) how much each value of X deviates from M; in other words, an index of dispersion would be the *Sum of (Individual Data Points—Mean of the Data Points)*.

Logical as this may seem, it has 2 drawbacks. The first difficulty is that the answer will be zero—not just in this situation, but in every case. By definition, the sum of the values above the mean is always equal to the sum of the values below it, and thus they'll cancel each other out. We can get around this problem by taking the absolute value of each difference (that is, we can ignore the sign whenever it's negative), but for a number of arcane reasons, statisticians don't like to use absolute numbers. Another way to eliminate negative values is to square them, since the square of any number—negative or positive—is always positive. So, what we now have is $\Sigma(X_i - M)^2$.

The second problem is that the result of this equation will increase as we add more subjects. Let's imagine that we have a sample of 25 values, with an SD of 10. If we now add another 25 subjects who look exactly the same, it makes intuitive sense that the dispersion of these 50 points should stay the same. Yet the formula as it now reads can result only in a larger sum as we add more data points. We can compensate for this by dividing by the number of subjects, N, so that the equation now reads $\Sigma(X_i - M)^2/N$.

In the true spirit of Murphy's Law, what we've done in solving these 2 difficulties is to create 2 new ones. The first (or should we say third, so we can keep track of our problems) is that now we are expressing the deviation in squared units; that is, if we were measuring IQs in children with autism, for instance, we may find that their mean IQ is 75 and their dispersion is 100 squared IQ points. But what in heaven's name is a squared IQ point? At least this problem is easy to cure: we simply take the square root of the answer, and we'll end up with a number that is in the original units of measurement, so in this example, the dispersion will be 10 IQ points, which is much easier to understand.

The last problem (yes, it really is the last one) is that the results of the formula as it exists so far produce a *biased* estimate, that is, one that is consistently either higher or (as in this case) lower than the "true" value. The explanation of this is a bit more complicated and requires somewhat of a detour. Most of the time when we do research, we are not interested so much in the samples we study as in the populations they come from. That is, if we look at the level of expressed emotion (EE) in the families of young schizophrenic males, our interest is in the families of all people who meet the criteria (the population), not just those in our study. What we do is

estimate the population mean and SD from our sample. Because all we are studying is a sample, however, these estimates will deviate by some unknown amount from the population values. In calculating the SD, we would ideally see how much each person's score deviates from the population mean, but all we have available to us is the sample mean. By definition, scores deviate less from their own mean than from any other number. So, when we do the calculation and subtract each score from the sample mean, the result will be smaller than if we subtracted each score from the population mean (which we don't know); hence, the result is biased downwards. To correct for this, we divide by N − 1 instead of N. Putting all of this together, we finally arrive at the formula for the standard deviation, which is:

$$SD = \sqrt{\frac{\Sigma(X_i - M)^2}{N - 1}}$$

(By the way, don't use this equation if, for whatever bizarre reason, you want to calculate the SD by hand, because it leads to too much rounding error. There is another formula, mathematically equivalent and found in any statistics book, which yields a more precise figure.)

Now that we've gone through all this work, what does it all mean? If we assume that the data are normally distributed, then knowing the mean and SD tells us everything we need to know about the distribution of scores. In any normal distribution, roughly two-thirds (actually, 68.2%) of the scores fall between −1 and +1 SD, and 95.4% between −2 and +2 SD. For example, most of the tests used for admission to graduate or professional schools (the GRE, MCAT, LSAT, and other instruments of torture) were originally designed to have a mean of 500 and an SD of 100. That means that 68% of people get scores between 400 and 600, and just over 95% between 300 and 700. Using a table of the normal curve (found in most statistics books), we can figure out exactly what proportion of people get scores above or below any given value. Conversely, if we want to fail the lowest 5% of test takers (as is done with the LMCCs), then knowing the mean and SD of this year's class and armed with the table, we can work out what the cut-off point should be.

So, to summarize, the SD tells us the distribution of *individual scores* around the mean. Now, let's turn our attention to the standard error.

Table I. Levels of LDE in the CSF of Administrators and Controls

Group	Number	Mean	SD
Administrators	25	25.83	5.72
Controls	25	17.25	4.36

Standard Error

I mentioned previously that the purpose of most studies is to estimate some population parameter, such as the mean, the SD, a correlation, or a proportion. Once we have that estimate, another question then arises: How accurate is our estimate? This may seem an unanswerable question; if we don't know what the

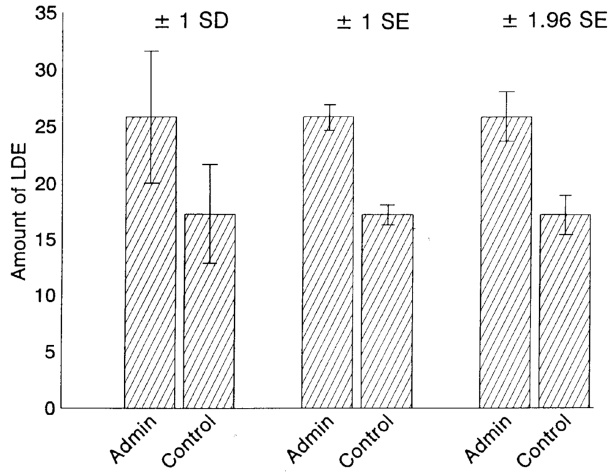


Figure 1. Data from Table I, plotted with different types of error bars.

population value is, how can we evaluate how close we are to it? Mere logic, however, has never stopped statisticians in the past, and it won't stop us now. What we can do is resort to probabilities: What is the probability (*P*) that the true (population) mean falls within a certain range of values? (To cite one of our mottos, "Statistics means you never have to say you're certain.")

One way to answer the question is to repeat the study a few hundred times, which will give us many estimates of the mean. We can then take the mean of these means, as well as figure out what the distribution of means is; that is, we can get the standard deviation of the mean values. Then, using the same table of the normal curve that we used previously, we can estimate what range of values would encompass 90% or 95% of the means. If each sample had been drawn from the population at random, we would be fairly safe in concluding that the true mean also falls within this range 90% or 95% of the time. We assign a new name to the standard deviation of the means: we call it the *standard error of the mean* (abbreviated as SEM, or, if there is no ambiguity that we're talking about the mean, SE).

But first, let's deal with one slight problem—replicating the study a few hundred times. Nowadays, it's hard enough to get money to do a study once, much less replicate it this many times (even assuming you would actually want to spend the rest of your life doing the same study over and over). Ever helpful, statisticians have figured out a way to determine the SE based on the results of a single study. Let's approach this first from an intuitive standpoint: What would make us more or less confident that our estimate of the population mean, based on our study, is accurate? One obvious thing would be the size of the study; the larger the sample size, N, the less chance that one or two aberrant values are distorting the results and the more likely it is that our estimate is close to the true value. So, some index of N should be in the denominator of SE, since the larger N is, the smaller SE would become. Second, and for similar reasons, the smaller the variability in the data, the more confident we are that one value (the mean) accurately reflects them. Thus, the SD

should be in the numerator: the larger it is, the larger SE will be, and we end up with the equation:

$$SE = \frac{SD}{\sqrt{N}}$$

(Why does the denominator read $\sqrt{N}$ instead of just N? Because we are really dividing the variance, which is SD^2, by N, but we end up again with squared units, so we take the square root of everything. Aren't you sorry you asked?)

So, the SD reflects the variability of *individual data points*, and the SE is the variability of *means*.

Confidence Intervals

In the previous section, on the SE, we spoke of a range of values in which we were 95% or 99% confident that the true value of the mean fell. Not surprisingly, this range is called the confidence interval, or CI. Let's see how it's calculated. If we turn again to our table of the normal curve, we'll find that 95% of the area falls between −1.96 and +1.96 SDs. Going back to our example of GREs and MCATs, which have a mean of 500 and an SD of 100, 95% of scores fall between 304 and 696. How did we get those figures? First, we multiplied the SD by 1.96, subtracted it from the mean to find the lower bound, and added it to the mean for the upper bound. The CI is calculated in the same way, except that we use the SE instead of the SD. So, the 95% CI is:

$$95\%\ \ CI = M \pm (1.96 \times SE)$$

For the 90% CI, we would use the value 1.65 instead of 1.96, and for the 99% CI, 2.58. Using the data from Table I, the SE for administrators is $5.72/\sqrt{25}$, or 1.14, and thus the 95% CI would be $25.83 \pm (1.96 \times 1.14)$, or 23.59 to 28.07. We would interpret this to mean that we are 95% confident that the value of LDE in the population of administrators is somewhere within this interval. If we wanted to be more confident, we would multiply 1.14 by 2.58; the penalty we pay for our increased confidence is a wider CI, so that we are less sure of the exact value.

The Choice of Units

Now we have the SD, the SE, and any one of a number of CIs, and the question becomes, which do we use, and when? Obviously, when we are describing the results of any study we've done, it is imperative that we report the SD. Just as obviously, armed with this and the sample size, it is a simple matter for the reader to figure out the SE and any CI. Do we gain anything by adding them? The answer, as usual, is yes and no.

Essentially, we want to convey to the reader that there will always be sample-to-sample variation and that the answers we get from one study wouldn't be exactly the same if the study were replicated. What we would like to show is how much of a difference in findings we can expect: just a few points either way, but not enough to substantially alter our conclusions, or so much that the next study is as likely to show

results going in the opposite direction as to replicate the findings. To some degree, this is what significance testing does—the lower the *P* level, the less likely the results are due simply to chance and the greater the probability that they will be repeated the next time around. Significance tests, however, are usually interpreted in an all-or-nothing manner: either the result was statistically significant or it wasn't, and a difference between group means that just barely squeaked under the $P < 0.05$ wire is often given as much credence as one that is highly unlikely to be due to chance.

If we used CIs, either in a table or a graph, it would be much easier for the reader to determine how much variation in results to expect from sample to sample. But which CI should we use? We could draw the error bars on a graph or show in a table a CI that is equal to exactly one SE. This has the advantages that we don't have to choose between the SE or the CI (they're identical) and that not much calculation is involved. Unfortunately, this choice of an interval conveys very little useful information. An error bar of plus and minus one SE is the same as the 68% CI; we would be 68% sure that the true mean (or difference between 2 means) fell within this range. The trouble is, we're more used to being 95% or 99% sure, not 68%. So, to begin with, let's forget about showing the SE: it tells us little that is useful, and its sole purpose is in calculating CIs.

What about the advice to use plus and minus 2 SEs in the graph? This makes more sense; 2 is a good approximation of 1.96, at least to the degree that graphics programs can display the value and our eyes discern it. The advantages are twofold. First, this method shows the 95% CI, which is more meaningful than 68%. Second, it allows us to do an "eyeball" test of significance, at least in the 2-group situation. If the top of the lower bar (the controls in Figure 1) and the bottom of the higher bar (the administrators) do not overlap, then the difference between the groups is significant at the 5% level or better. Thus we would say that, in this example, the 2 groups were significantly different from one another. If we actually did a *t* test, we would find this to be true: $t(48) = 2.668$, $P < 0.05$. This doesn't work too accurately if there are more than 2 groups, since we have the issue of multiple tests to deal with (for example, Group 1 versus Group 2, Group 2 versus 3, and Group 1 versus 3), but it gives a rough indication of where the differences lie. Needless to say, when presenting the CI in a table, you should give the exact values (multiply by 1.96, not 2).

Wrapping Up

The SD indicates the dispersion of individual data values around their mean, and should be given any time we report data. The SE is an index of the variability of the means that would be expected if the study were exactly replicated a large number of times. By itself, this measure doesn't convey much useful information. Its main function is to help construct 95% and 99% CIs, which can supplement statistical significance testing and indicate the range within which the true mean or difference between means may be found. Some journals have dropped significance testing entirely and replaced it with the reporting of CIs; this is probably going too far, since both have advantages, and both can be misused to equal degrees. For example, a study using a small sample size may report that the difference between the control

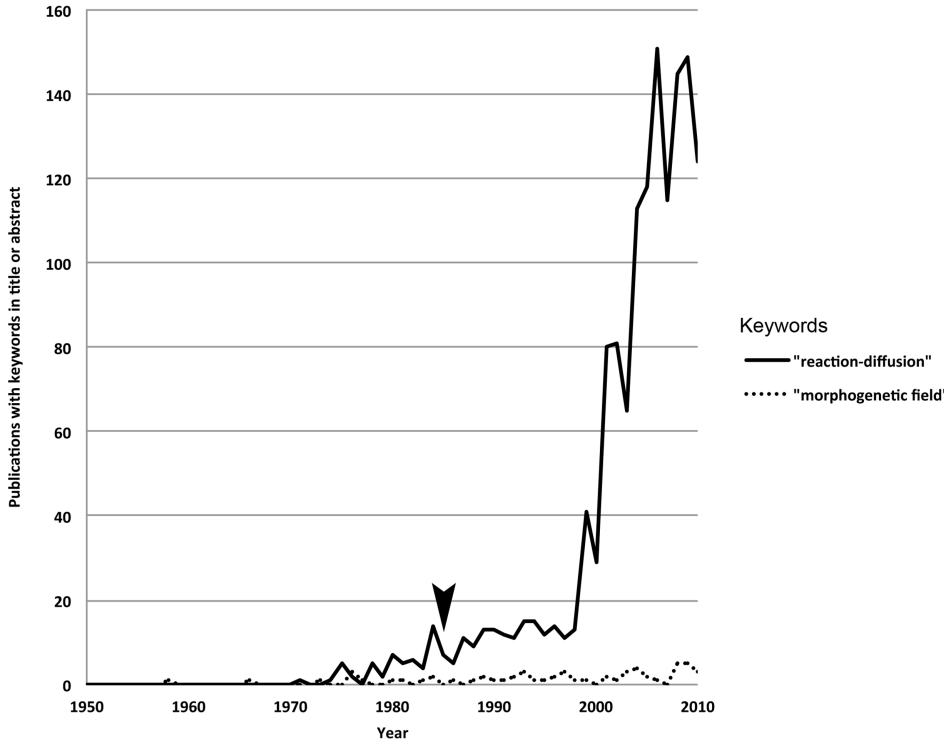
and experimental group is significant at the 0.05 level. Had the study indicated the CIs, however, it would be more apparent to the reader that the CI is very wide and the estimate of the difference is crude, at best. By contrast, the much-touted figure of the number of people affected by second-hand smoke is actually not the estimate of the mean. The best estimate of the mean is zero, and it has a very broad CI; what is reported is the upper end of that CI.

To sum up, SDs, significance testing, and 95% or 99% CIs should be reported to help the reader; all are informative and complement, rather than replace, each other. Conversely, "naked" SEs don't tell us much by themselves, and more or less just take up space in a report. Conducting our studies with these guidelines in mind may help us to maintain the standards in psychiatric research.

Résumé

Beaucoup de gens confondent l'écart-type et l'erreur-type de la moyenne et ne savent pas lequel utiliser pour présenter les données sous forme graphique ou tabulaire. L'ecart-type indique la variabilité des données originales et devrait être mentionné pour toutes les etudes. L'erreur-type montre la variabilité des valeurs moyennes, comme si l'étude avait été reprise de nombreuses fois. En soi, l'erreur-type n'a pas d'utilité particulière; toutefois, on s'en sert pour créer les intervalles de confiance à 95% et à 99% utilises pour établir la fourchette de valeurs dans laquelle se situe la valeur «réelle». Les intervalles de confiance signalent au lecteur la précision des estimations des valeurs démographiques. Lorsqu'on se sert de graphiques, la barre d'erreur représente un intervalle de plus à moins 2 écarts-types (ce qui correspond à l'intervalle de confiance de 95%). Elle devrait entourer la valeur moyenne. Les épreuves de signification statistique et les intervalles de confiance présentent une grande utilité, car ils aident le lecteur à établir l'importance des constatations.

Case Study 16: Field Models of Pattern Formation



Goodwin (1985) Developing organisms as self-organizing fields. In this graphic we have compared this theoretical approach to the frequently cited reaction-diffusion theory.

Perspective on *Developing Organisms as Self-Organizing Fields*

Wendy Brandts, PhD
University of Ottawa,

Understanding complex systems top-down

I. Introduction

Against the tide of fashion dictating that biological systems are to be understood primarily from their molecular- and cellular-level behaviours, this ground-breaking paper makes the case that understanding biological systems requires structural laws at multiple levels. Drawing on the deep biological understanding and unique perspective of Goodwin, and the insight and knowledge of quantitative structural laws of L.E.H. Trainor, a mathematical framework is proposed for structural laws of biological form.

After all, biological systems are also physical systems, with some special additional properties. The historical success of physics to describe physical systems (e.g. quarks, nucleons, atoms, molecules) via structural laws at many levels suggests carrying out a similar program at higher levels of organization and complexity.

II. Structuralist Approach

Goodwin takes the structuralist point of view, going back to Hans Driesch and Darcy Thompson, that seek laws of form showing regularities of morphology over many taxonomic groups. He builds on the approach of the Clockface Model (French *et al* 1976) which are rules of morphogenesis making no reference to composition or inheritance. In particular, Goodwin's morphogenetic field serves to inform different parts of an organism 'where in the whole it is'.

In this interpretation, the morphogenetic field is the self-organizing entity. The organism, not the cell, is the fundamental biological entity. Spatial organization of the whole derives from the global morphogenetic field plus the constraints of the parts.

Goodwin and Trainor (1980) provide a mathematical formalism for this morphogenetic field, which was not attained by the creators of the Clockface Model. The field is a structure: it belongs to an invariant set defined by internal relations (the field equations). An analogy from classical mechanics is the trajectories of motion of an object in a central force field, where the set of possible solutions to Newton'ss laws are restricted to the conic sections.

The formalism of this morphogenetic field is akin to electrodynamic field equations of physics. The concept of 'selection rules', also borrowed from physics, determines which possible solutions of the field equations will actually be manifested in the system, subject to other constraints. The resemblance to physical theory in this approach echoes the background of the late L.E.H. Trainor, originally a nuclear physicist, who developed the new field of theoretical biophysics at the University of

Toronto with inspiration from Goodwin, C.J. Lumsden and others (see various authors, [Physical Theory in Biology: Foundations and Explorations (1997)]).

The use of physical fields is both convenient and natural because the proposed biological fields actually demonstrate properties similar to physical fields, e.g. smoothness and continuity (one cell is more similar to a neighbouring cell than to a distant cell). The holistic nature of physical fields which allows, for example, a unique electric field to be determined everywhere in a region given sufficient information about the field on the boundary, is reminiscent of the ability of a morphogenetic field to interpolate in regeneration, and thus to reform a whole organism from the information contained in some of the parts.

With this formulation of the morphogenetic field, Goodwin describes early cleavage processes by applying the Laplace equation over a sphere (representing the blastula) and selecting solutions of spherical harmonics whose null lines then indicate cleavage planes.

He concludes that constraints determining specific forms arise from other organizational levels than simple molecular composition and therefore top-down models are required for understanding. Gene products act within the context of fields (electric, visco-elastic, etc) which generate morphology. He argues the inadequacy of any theory of biological form based only on genes and molecules.

III. Models based on Goodwin's Morphogenetic Field Approach.

In the years since Goodwin's pioneering papers, the morphogenetic field formalism (Goodwin and Trainor 1980) has formed the starting point for a number of more detailed applications, which in turn confirm the usefulness of the fundamental concepts.

1. Limb regeneration in amphibians

Valuable predictions can be made, such as the outcome of transplant experiments, without needing to know all the underlying details of structure and composition. For example, a field model of limb regeneration based on fields with solutions of specific symmetry can predict handedness of supernumerary growth (Tevlin and Trainor 1985, Totafurno and Trainor 1987).

2.Regulation in a single cell

Analogous to multi-cellular pattern formation, the growth is governed by global patterning fields, not based on genes or molecules. The field model describes the organism at a 'high level' or macroscopic level of the structural hierarchy, organizing large and various collections of experimental data and predicting developmental configurations and pathways (Brandts and Trainor 1990a, Brandts and Trainor 1990b, Brandts 1993).

3. Electrical signals (dynamics of voltage potentials in gap junction-coupled bioelectric network within non-neural tissues) as global patterning fields (Levin *et al* 2015).

These applications all demonstrate the power of top-down models.

While morphogenetic field models stand on their own as useful tools, it may nevertheless be informative to investigate what underlies the morphogenetic field. Many different lower-level factors could contribute to the expression and maintenance of the field.

Some examples are: order parameters relating to the organization of microfilaments in the cell surface (Goodwin 1980); viscoelastic forces (Hart *et al* 1989); reaction-diffusion of a binary fluid (Brandts 1995); electrical potentials (Levin *et al* 2015).

IV. Recent ideas that support Goodwin's perspective.

The structuralist approach embraced by Goodwin has been largely underappreciated due to fashions and successes in microbiology. Scientific progress in recent years in the new field of complexity has brought several fundamental issues to the fore that should encourage a more general acceptance of Goodwin's approach. Here I touch on three of these issues:

1. Simulation *vs* duplication

A simulation of the behavior of a system does not necessarily duplicate the way a system itself actually operates. Nevertheless, simulating or modeling the behaviour of a complex system through top-down processes can aid understanding. After all, no one questions whether a particle calculates its velocity while deciding how to move according to laws of motion; the laws are valuable for predicting behavior, and making control decisions, regardless of the interpretation.

Complex system behaviour can often be viewed as dynamics or as computation. [Brandts (1997). 'Complexity: A Pluralistic Perspective'.].

The point is, when systems behave 'as if' they were calculating, they can fruitfully be modelled as such. Regardless of the philosophical interpretations, such as those entertained by Einstein and Bohr on Quantum Mechanics, in the end, the predictive power of a model wins out.

Recent concerns about goal-directed behaviour and the apparent role of teleology in top-down models may thus be dismissed.

2. Emergence and reducibility

One definition of emergence (from physics) is that emergence is a new behaviour that arises at a higher level and which can not be predicted from the lower level behavoirs. For example, in physics the exchange symmetry of fermions, which influences the structure of the periodic table, is a rule which applies to multi-fermion systems but cannot be derived from the behaviour of a single fermion. See also the two proposed broad and distinctive classes of emergence, *reducible* and *a prior irreducible* (Trainor 1997, Trainor 1985).

In a different case, consider Thermodynamics, TD, the higher-level macroscopic model of the behaviour of the gaseous state of matter. The concepts and laws of TD relate observable quantities, namely temperature and pressure, and provided

essential understanding to power the industrial revolution. But TD laws do not apply to single molecules, or even require any particular molecular structure. The lower-level model, Statistical Mechanics, which was developed later, provides the microscopic theoretical underpinnings of TD; nevertheless, TD concepts and laws are still useful and applicable today.

In addition, it is important to discriminate between reduction of *theories* and reduction of *phenomenon* [Brandts (1997). 'Complexity: A Pluralistic Perspective'.]. If the high level processes are irreducible, then they will certainly not be explicable by a lower-level theory—i.e. a higher level or top-down theory will be *essential* for understanding. But even if the *processes* are reducible, the *theories* may not be.

Moreover, while the biological distinction between local and global interactions can often be made clear, and tested in a laboratory situation, in the mathematics of modeling, the distinction between global or local control can be ambiguous. This is because dynamical systems often have alternative mathematical representations which may be entirely equivalent but lead to seemingly different interpretations. An example from classical physics is mechanics, in which a particle can be described as moving according to Newton's laws, which are local, or so as to minimize action over any segment of its future path, which is non-local in both space and time.

In the particular case of field models, the mathematical expression of pattern control is formulated in terms of minimization of total energy, a global property. However, the model could in principle have been formulated in terms of differential equations which express field values and interactions locally. In fact, there are many different dynamical processes (equations) which could give rise to the same steady state minimum energy solutions of the global field.

3. Complexity

Complexity exists on many levels and scales in biological systems. Even defining a quantitative measure of complexity is a daunting task, demanding that many different levels of explanation be employed.

Recent developments in the science of complexity and consideration of biological systems as complex systems has lead to the realization multiple levels of modeling will be required to understand complex biological systems (Brandts 1997). Top-down modeling is essential for a pluralistic approach; no single level of analysis has priority. Each description or model naturally emphasizes its own variables and should be considered a legitimate method of modeling without requiring descriptions and their language to translate into each other in a reductionist fashion (although it may sometimes be possible to do so).

Laws and models are simplifications of the world. For example, even position and momentum, in Newton's classical mechanics, are properties of the formalism, and do not necessarily reflect the real world because position and momentum do not exist simultaneously in the lower-level Quantum Mechanical formulation. That is, Newton's formalism is only applicable to the real world over a restricted range of velocities, timescales, and sizes; at other levels, new phenomenology and models are required. Real systems map to many formal systems, each one capturing different aspects of behaviour. There are distinctive ideas in biology not translatable into physics.

In complex systems such as biological ones, incorporating all lower-level details is unwieldy or impossible, and top-down models like the field approach are required in order to make predictions about higher-level biological behaviours. High level models, with all their simplifications, are needed because when models contain 'everything'—field data, a plethora of parameters, a patchwork of dynamics—and simulations are run as a black box, little in the way of strategic relationships between variables, or the effect of the numerous parameters on the dynamics, can be uncovered (Brandts 2002). We may simulate nature, but we will gain little insight, or control.

Or as J.L. Borges put it, *Of what use is a map as big as the world?*

References

Brandts W A M and Trainor L E H 1990a A nonlinear field model of pattern formation -- intercalation in morphalactic regulation *J. Theor. Biol.* **146** 37–57

Brandts W A M and Trainor L E H 1990b A nonlinear field model of pattern formation -- application to intracellular pattern reversal in tetrahymena *J. Theor. Biol.* **146** 57–87

Brandts W A 1993 A field model of left-right asymmetries in the pattern regulation of a cell *IMA J. Math. Appl. Med. Biol.* **10** 31–50

Brandts W A M 1995 Relevance of field models to global patterning in ciliates *Interplay of Genetic and Physical Processes in the Development of Biological Form* ed E Beysens, G Forgacs and F Gaill (Singapore: World Scientific)

Brandts W A 1997 Complexity: A Pluralistic Perspective *Physical Theory in Biology: Foundations and Explorations* ed C J Lumsden, W A Brandts and L E H Trainor (London: World Scientific Publishing)

Brandts and others 1997 *Physical Theory in Biology: Foundations and Explorations* ed C J Lumsden, W A Brandts and L E H Trainor (London: World Scientific Publishing)

Brandts W A M 2002 Tetrahymena and Ants -- Simple Models of Complex Systems *Formal Descriptions of Developing Systems. Nato Science Series*v. 121 ed J Nation, I Trofimova, J D Rand and W Sulis (Kluwer Academic Publishers)

French V, Bryant P J and Bryant S V 1976 Pattern Regulation in Epimorphic Fields *Science* **193** 969–81

Goodwin B C and Trainor L E H 1980 A field description of the cleavage process in embryogenesis *J. Theor. Biol.* **85** 757–70

Levin M, Pezzulo G and Levin M 2015 Re-membering the body: applications of computational neuroscience to the top-down control of regeneration of limbs and other complex organs *Integrative Biology* **7** 1487–517

Tevlin P and Trainor L E H 1985 A two vector field model of limb regeneration and transplant phenomena *J. Theor. Biol.* **115** 495–513

Totafurno J and Trainor L E H 1987 A non-linear vector field model of supernumerary limb production in salamanders *J. Theor. Biol.* 124

Trainor L E H 1997 *Physical Theory in Biology: Foundations and Explorations* ed C J Lumsden, W A Brandts and L E H Trainor (London: World Scientific Publishing)

Trainor L E H 1985 Remarks on emergence in physics and biology *Mathematical Essays on Growth and the Emergence of Form* ed P Antonelli (Edmonton: University of Alberta Press)

DEVELOPING ORGANISMS AS SELF-ORGANIZING FIELDS

B.C. *GOODWIN*

5.1 Introduction

Self-organization is a property of a structurally and functionally integrated entity which is considered to be made up of, or to have, "parts." There are essentially two ways of relating the whole and the parts: either the parts are regarded as given (with or without some temporal sequence of presentation or production) and the entity is generated by their interaction; or the entity is regarded as a whole defined by certain invariant relations, the "parts" coming into being as a result of systematic transformations which preserve invariance while generating heterogeneity ("parts") within a functional and structural unity. The most extreme form of the first description is atomism, which assumes that all the information necessary for generating the entity is resident in the parts, so that spontaneous assembly occurs simply as a result of their interaction. This hypothesis, in various forms and with various modifications, is a dominant theme in contemporary models of evolutionary, developmental, behavioral, and even cognitive processes. What characterizes this conceptualization is the assumption that there are no laws of self- organization other than those governing the interaction of the parts or constituents so that the whole is reducible to these parts and their interactions. In developmental biology, the constituents are usually considered to be, ultimately, molecules, although there are theories in which cells are used as intermediate "atoms" in the analysis. Since genes are generally considered to be the determinants of which molecules are present in organisms, it follows that organisms are reducible to genes. The self-organizing process of embryo-genesis is then regarded as a consequence of two activities: the operation of the "genetic program" which determines the types of molecular constituents in the organism, the sequence and spatial location of their appearance; and the interaction of these constituents according to physical and chemical laws.

Such an approach to embryogenesis means that the specific forms generated by morphogenetic processes, defining different species of organism, are irreducibly complex because there are no laws or principles which constrain the "genetic program" which is the determinant of specific form. This program is the result of random permutation and natural selection, purely contingent processes as far as organisms are concerned, the only constraint being that the organism specified by the program must be able to survive and leave offspring in some environment. In such a view, there can be no general laws of biological form. Each species is, speaking more than metaphorically, a law unto itself. This is reflected in the primacy of the species concept in neo-Darwinism and the emphasis placed upon competition and survival as the expression of the individual species' success in establishing a unique, singular relationship of order and stability within itself and with its environments. Thus, as one might expect since cognitive constructs are themselves

"self- j organizing," there is a clearly-defined continuity between the way organisms are conceptualized in neo-Darwinism as "survival machines" and the way in which they are considered to be generated from their molecular parts.

There is no *a priori* reason why such a description of self-organization should not be valid, and indeed it is clear that there are special cases in both the animate and the inanimate realms where atomistic explanations appear to be appropriate. As regards structure or form, these are the instances where a crystallization or "self-assembly" type of process leads to a unique morphology. However, even in inorganic chemisty one encounters many instances of polymorphism in which the same substance can crystallize into different forms, familiar examples being graphite and diamond, or rhombic and mono-clinic sulphur. Thus, in general, composition does not determine form. A,similar polymorphism is observed in biological structures at different levels of organization, the molecular (Oosawa et al., 1966), the cellular (Sonneborn, 1970), the tissue (Saunders et al., 1957) and so on. Much of embryology consists in fact of generating a variety of "abnormal" forms out of cells of identical genotype: for example, the induction of supernumerary limbs in an amphibian by a simple manipulation of tissue in the embryonic limb bud involving no addition or deletion of cells, but simply a change of relative position, and no change in the external environment. Furthermore, organisms containing specific mutant genes (e.g. homoeotics) may or may not express them; while organisms with wild-type genes can show the "mutant" phenotype (spontaneous homoeotic transformation). Hence there is no one-to-one relationship between genotype and phenotype, always assuming j a constant external environment, so that genes are not the specific determinants of morphology. That is to say, the form of an organism is not determined by its genome, with the consequence that self-assembly theories together with a genetic program are inadequate to provide a generative theory of biological self-organization (see Webster and Goodwin, 1982, for a more detailed argument leading to this conclusion).

We must now consider whether or not there is empirical evidence relating to general organizational principles, or laws of form in biology, manifesting as regularities of morphology over large taxonomic groups. We have seen above that neo-Darwinism, which takes the view that "the chief part of the organization of every being is simply cue to its inheritance" (Darwin, 1859; inheritance meaning, in contemporary usage, the genetic program), provides no basis for understanding any such regularities since biological form in this theory is determined by contingency, not by law. However, if ordering constraints do exist in the biological realm, then this must be taken into account of in any theory of biological self-organization. This leads us to the work of the pre-Darwinian rational morphologists, who were animated by a belief in the possibility of a rational, intelligible ordering or classification of organisms which would provide an insight into the laws of organic creation (i.e. generative rules). This tradition reached its peak in the work and insights of the great comparative morphologists of the late eighteenth and early nineteenth centuries such as Geoffroy St. Hilaire, Cuvier, Reichert, and Owen, who

searched for and discovered empirical regularities of organismic structure. These regularities appeared as invariant structural relations or "typical forms" which were seen to define that which is common to a variety of particular realizations of the same type. Owen's demonstration of the structural homologies which exist between the great variety of vertebrate limbs, leading to the concept of the pentadactyl limbs as the typical form, is characteristic of this work. Each specific member of the invariant set, such as the limb of the horse, of the bat, of the frog, etc., can then be seen as equivalent to every other member under a transformation, so that a common plan is revealed which unifies the diversity of manifest forms. It is, in fact, straightforward to demonstrate the simple proposition that tetrapod limb morphogenesis may be understood in terms of some basic generative principles capable of producing a great variety of limb forms which are all transformable one into the other under modifications of the limb generating process (Goodwin and Trainor, 1982). This is analogous to the realization that the different forms of motion shown by bodies under the action of a central attracting force, obeying Newton's laws, all belong to the same invariant set known as the conic sections; and indeed the rational morphologists were inspired by the same vision as Newton, which was the Enlightenment Ideal of a mathematical natural science. Their conviction was, and they provided good evidence for the belief, that the morphological complexity of organisms is not irreducibility complex but that there exist rational principles or laws of form which render the diversity intelligible.

Despite the fact that this tradition was largely eclipsed by Darwinism, which adopted the diametrically opposite view that organismic form is determined not by rational law but by historical accident, by contingency, a few rather isolated and sometimes misunderstood biologists have pursued this approach further. Among these the embryologist Hans Driesch stands out and his work is very relevant to the view of self-organization which will be developed below. He introduced the field concept into embryology as a result of his demonstration that relative position in the whole embryo is an important determinant of cell fate. He used the concepts of wholeness, self-regulation, and transformation to define the properties of tissues which respond to a variety of disturbances (e.g., removal or addition of cells, or spatial reordering of parts) by a reorganization such that tire normal form is generated. Examples of such fields are the amphibian embryo from fertilization up to about the gastrula stage, the limb and eye primordia, and a variety of other tissues domains which define secondary fields. Within such domains, relative position is a primary determinant of cell fate and the parts which emerge during individuation and differentiation come into being as a result of local and global ordering principles, generating a structural and functional unity. Driesch assumed, like the rational morphologists, that there are organizing rules which operate within organisms to constrain or limit the forms which can be generated (Driesch, 1929) but, again like his predecessors, he failed to give them any mathematical formulation. The problem to be addressed now is what type of mathematical description may be appropriate for these organizing principles, for which there is clear biological evidence.

5.2 Organisms as Fields

The proposition which emerges from the above analysis is that living entities are wholes or *structures* defined by internal relations which remain invariant under certain categories of transformation, the latter limiting the possible generative processes which can result in organisms of specific form (species). Organisms are not, in this view, generated as a result of the interaction of "atomic" constituents, whatever these may be construed to be. Heterogeneity ("parts") arises as a result of systematic transformations of the organized whole, which may be described as the manifestation of states selected from a potential set which satisfies a primary property of invariance characteristic of organisms. Thus the organism is not so much a self- organizing system which generates an ordered state from disordered or less ordered parts; it is more a self-organized entity which can undergo transformations preserving this state. The problems faced by this conceptualization are those of making explicit the nature of the invariant internal relationships which define the whole; the type of transformation which it can undergo; and the relationship between whole and part which confers upon it the properties of generation (reproduction) and regeneration.

Following Driesch's insight that developing organisms have field properties, we may proceed to the question of what type of field and how it may be characterized mathematically. A very extensive body of experimental work in developmental biology since Driesch's pioneering studies has led recently to the observation that an appropriate arithmetic description of the spatial smoothness characteristic of developmental fields is a simple spatial averaging or intercalculation rule applied to field values (French, Bryant, and Bryant, 1976). This states simply that the field value at any point within the boundaries of a developmental field is the arithmetic mean of the values at equidistant neighboring points. Mathematically, this leads to the most general field equation used in physics, namely Laplace's equation. The question then naturally arises whether one can use solutions of this and related equations, known as harmonic functions, to describe developmental fields and hence biological form. Preliminary essays in this direction have been published (Goodwin, 1980; Goodwin and Trainor, 1980; Goodwin and Trainor, 1982). This approach will now be illustrated by an analysis of the earliest stage of amphibian embryogenesis, following the treatment of Goodwin and Trainor (1980), and then certain conclusions regarding the problem of self-organization in biology will be drawn.

5.3 A Field Description of the Typical Cleavage Process

The first five stages of the typical cleavage pattern is described by classical investigation as shown in Figure 5.3-I, starting from the two-cell stage after the first division of the egg. From the 32-cell stage, cell divisions continue to show an alternation between vertical and horizontal cleavage planes, but there is at some stage a loss of spatial and temporal order (synchrony) which differs between species. Since our interest is in the geometry of the cleavage planes, we project the typical pattern onto the original spherical egg to get the schematic sequence in Figure 5.3-II, which illustrates the first seven cleavages up to the 128-cell morula. This is immediately suggestive of a sequence of harmonic functions on the sphere, the

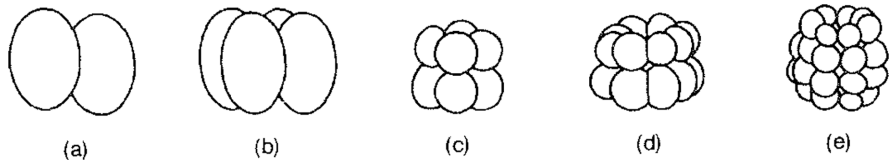


Figure 5.3-I. The holoblastic radial cleavage pattern.

cleavage lines corresponding to nodal lines of spherical harmonics. Accordingly, we develop an "eigenfunction" description of cleavage on the basis of a minimization principle wherein the eigenstates of a morphogenetic surface field describe the successive stages of the cleavage process in the early embryo. The cleavage process is then seen as a series of transformations to successively higher characteristic states of the morphogenetic field as metabolism proceeds. An analogy to this may be found in the electron density distributions of the hydrogen atom, in which the transitions to successively higher energy states results from the action of some external optical pumping field. A characteristic biological feature is that the cleavage transformations are "pumped" or induced internally, the system being self-generating.

5.4 A Variational Principle for Cleavage Planes

Proceeding with the analysis at a fairly abstract level, let us now introduce a field function $u(\theta, \phi)$ over the surface of the sphere and adopt the convention that its nodal lines represent lines of least resistance to a furrowing process preliminary to the development of cleavage planes. This function may be taken to be some kind of order-disorder parameter relating to the organization of microfilaments in the cell surface. The basic stability of the typical cleavage pattern suggests the use of a minimization principle on this field function. An appropriate surface density function, which in a physical problem would be the energy density, is

$$E(\theta, \phi) = A\left[\left(\frac{\partial u}{\partial \theta}\right)^2 + \frac{1}{\sin^2\theta}\left(\frac{\partial u}{\partial f}\right)^2 + \beta u^2\right] \tag{5.4-1}$$

where the constants A and β incorporate relevant physiological units. Then suppose that the characteristic cleavage planes correspond to a minimum of the integral of this density function over the surface energy E,

$$\delta \int_0^{2\pi} \int_0^{\pi} E(\theta, \phi) \sin\theta \, d\theta \, d\phi = 0 \tag{5.4-2}$$

subject to a conservation law on u^2

$$\int_0^{2\pi} \int_0^{\pi} u^2(\theta, \phi) \sin\theta \, d\theta \, d\phi = 1 \tag{5.4-3}$$

which amounts to a normalization condition on the field variable u. Equations (5.4-2) and (5.4-3) require satisfaction of the Euler-Lagrange equation (see Trainor and Wise, 1979)

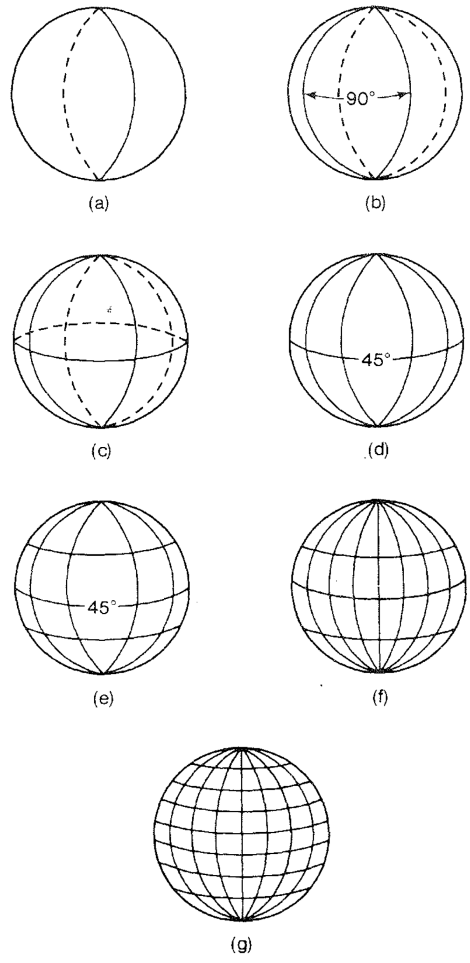


Figure 5.3-II. The geometry of typical cleavage planes up to the 128-cell stage.

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial^2 \mathrm{u}}{\partial\phi^e}-\alpha \mathrm{u}=0 \tag{5.4-4}$$

where α: incorporates the parameter β and an undetermined multiplier of equation (5.4-3).

The usual conditions on u, that it be finite, single-valued and continuous over the sphere restricts the possible solution u to a characteristic (eigen-function) set, viz. the spherical harmonics (real part taken):

$$u(\theta,\phi)\rightarrow Y_{\ell m}(\theta,\phi)=\sqrt{2}\,\mathrm{N}_{\ell \mathrm{m}}\mathrm{P}_\ell^{\mathrm{m}}(\cos\theta)\ \mathrm{m}\phi \tag{5.4-5}$$

where ℓ takes on the integral values 0, 1, 2, etc., and for given ℓ, the m values are integers ranging from $-\ell$ to $+\ell$. The parameter α is restricted to the corresponding characteristic values (eigenvalues) $\ell\,(\ell+1)$. In equation (5.4-5), the $\mathrm{P}_\ell^{\mathrm{m}}$ are associated

Legendre polynomials (Hobson, 1955) and the $N_{\ell m}$ are the normalization constants given by

$$N_{\ell m} = \left[\frac{(2\ell + 1)(\ell - |m|!}{4\pi(\ell + |m|)!} \right]^{1/2} \tag{5.4-6}$$

It is easy to calculate the surface "energy" corresponding to each characteristic cleavage state. The result is:

$$E_\ell = A\{\ell(\ell + 1) + \beta\}. \tag{5.4-7}$$

However, not every characteristic state (5.4-5) is realized in the cleavage process since the mitotic apparatus imposes a biological constraint (somewhat analogous to superselection rules in physics: Wick, Wightman and Wigner, 1952) that the number of cells is doubled in each cleavage stage.

According to the ideas set out above, the nodal lines of the characteristic function $Y_{\ell m}(\theta, \phi)$ on the sphere are in correspondence with the furrow lines of a characteristic cleavage state, except that the set of characteristic states is limited by the requirement that the number of cells is given by 2^p where p is the number of cell divisions. It is easily shown that the number of cells in a state characterized by $Y_{\ell m}$ is $2m(\ell - m+l)$ unless $m = 0$ in which case the number is $\ell + 1$. (It is sufficient to choose the real part of $Y_{\ell m}$, i.e. the cos $m\phi$ solutions, so that we need consider only in $m \geqslant 0$.) Hence the biological constraint requires that

$$\begin{aligned} 2^p = 2m(\ell - m + 1) \quad &\text{if} \quad m \neq 0, \quad \text{or} \\ = \ell + 1 \qquad\quad &\text{if} \qquad m = 0. \end{aligned} \tag{5.4-8}$$

In general this equation, for a given characteristic cleavage state corresponding to p divisions, is satisfied by more than one set of (ℓ, m) values. It is natural to suppose that the choice is made primarily on the basis of lowest ℓ value, since according to equation (5.4-7), this minimizes the "energy." The choice is then nearly unique, except for a two-fold degeneracy every second division. It is assumed that the animal-vegetal polarity of the embryo defines a secondary polar field weaker than the primary field which removes this degeneracy in favor of the highest m value for a given ℓ, in much the same way as a magnetic field removes the $(2\ell + 1)$ fold degeneracy of magnetic- states in the hydrogen atom.

Table 5.4-1 shows the correspondence between cleavage states and characteristic functions $Y_{\ell m} = 2N_{\ell m} P_\ell^m \cos m\phi$ by listing the number of cells to be expected from each set of (ℓ, m) values up to $\ell = 7$. The appropriate (ℓ, m) pair is then selected out uniquely by the conditions expressed in equations (5.4-8), together with the minimization condition (5.4-7), except for the two-fold degeneracies at the first, second, fourth, etc., cell divisions. As remarked above a unique correspondence is achieved by assuming that a weaker polar field selects (1,1) over (1,0), (2,2) over (2,1) and (5,4) over (5,2), that is, it favors highest m value for a given ℓ.

In Table 5.4-1 the selected states for the 6th and 7th cleavages corresponding to the 64 and 128 cell stages have been included, without listing all of the rejected (ℓ, m) values. Figure 5.3-II illustrates the nodal lines for the successive stages in a typical cleavage process. (Note that in the model used here the furrowing process

corresponds to an accumulative conjunction of surface harmonic patterns. In a variant of the model (Goodwin and LaCroix, 1982), the pattern is not cumulative and each stage is described by a single harmonic pattern of nodal lines.)

Table 5.4-1. Correspondence between $Y_{\ell m}(\theta, \phi)$ and cell number, defined by $\ell + 1$ if m = 0, otherwise by $2m(\ell - m + l)$.

ℓ value	m value	cell number	(ℓ, m) pair selected
1	0	2	(1, 1)
	1	2	
	0	3	
2	1	4	(2, 2)
	2	4	
	0	4	
3	1	6	(3, 2)
	2	8	
	3	6	
	0	5	
	1	8	
4	2	12	
	3	12	
	4	8	
	0	6	
	1	10	
5	2	16	(5, 4)
	3	18	
	4	16	
	5	10	
	0	7	
	1	12	
	2	20	
6	3	24	
	4	24	
	5	20	
	6	12	
	0	8	
	1	14	
	2	24	
7	3	30	(7,4)
	4	32	
	5	30	
	6	24	
	7	14	
11	8	64	(11,8)
15	8	128	(15, 8)

5.5 Fields and Self-Organization

This example clarifies some of the abstract concepts introduced in the discussion on organisms as fields, and it is now of some interest to elaborate on these and their relationship to the concept of self-organization in biology. The first point to emerge is the primacy of the organism as the fundamental biological entity, replacing the usual definition of the cell as the unit of life. This follows from the fact that the field is the self-organizing entity, and this is coextensive in the above description with the organism. The orderly geometry of the cleavage planes is a reflection of this organization at the level of the developing embryo. This does not mean that the parts, which in this case are cells, have no properties of their own. On the contrary, the constraint of binary division, a cellular property, is a major source of the distinctive geometry of the cleavage process, since this defines one of the three selection rules which determine the harmonic functions allowable as descriptions of this process. What thus becomes apparent is that the spatial organization of the whole derives from principles relating to global field behavior together with constraints coming from the properties of the entities which are generated as parts. Such a description avoids both atomistic reduction and a holistic description which identifies the whole with some (conceptually or materially) isolable essence. The genome is such an isolable essence, traceable historically to its conceptual roots in idealistic holism via Weismann's conceptualization of the organism as separable into germ plasm (essence) and somatoplasm (expression of the essence; see Webster and Goodwin, 1982). A field description of developing organisms sees spatial organization as the expression of distributed influence, global order being constrained to give specific morphology as a result of autonomous properties of parts. This defines a "decentered structure."

The ambiguity between the concepts of "cell" and "organism" can be resolved in terms of the above description. The fertilized egg is both a cell and a developing organism. It is an organism insofar as it is a totality describable by a field; it is a cell insofar as it embodies the specific constraints (e.g., binary division) characteristic of such an entity. As cleavage proceeds, the organism continues to be identified with the whole field (the embryo), while cells are identified as parts which play specific roles within a transforming context. After gastrulation, more complex parts such as neural plate, limb fields, eye fields, etc., come into existence. These consist of aggregates of cells so that an integrated hierarchy of parts emerges within the context of the organism as the global field which continues to impose organizational constraints upon the whole, the parts imposing reciprocal constraints so that further specific form emerges.

This type of description allows one to make comparisons between developmental processes in a cellular (or unicellular) and in multicellular organisms and to understand them in the same terms, which is the classical view. If one adopts a position such as that of Wolpert (1971), that development is to be understood in terms of the responses of cells (really, of their genomes: Wolpert and Lewis, 1975) to "positional information" established over multicellular domains, then such comparisons become problematical, as discussed by Frankel (1974). The view of organisms as fields overcomes this difficulty and leads to the suggestion of some unexpected

homologies between the morphology of ciliate protozoa and of amphibian gastrulae (Goodwin, 1980). This is very much in the tradition of rational morphology, since this makes clear that one must seek homologies at the level of "deep structure," i.e. at the level of generative principles such as those defining field properties, not in terms of "surface structure" such as whether or not an organism is partitioned into cells. Again, this is not to deny that surface structure imposes constraints which affect manifest form; it is simply that in comparing field effects at the level of the whole organism, these are secondary.

A very important problem in embryogenesis relates to the means whereby new or emergent aspects of morphogenesis are initiated at specific times in the process. For example, the relatively simple recursive process of cleavage in amphibian embryos is followed by the dramatic phenomenon of gastrulation, which transforms the hollow blastula into the three-layered late gastrula. The transition from cleavage to gastrulation has been described (Goodwin, 1980) as the expression of a cortical or surface field which is radially symmetric in the unfertilized egg but develops bilateral symmetry, normally as a result of sperm entry. The description of such a surface field in terms of harmonic functions reveals that with bilateral symmetry there appears a special point on the surface, a saddle point, which is identified with the future dorsal lip of the blastopore. However, it is assumed that the influence of this singularity on the morphogenetic process cannot become significant and be expressed until cleavage transforms the initially solid sphere of the egg into a hollow spherical shell, the blastula. Then the saddle point can make its presence felt as a point on the blastula where surface polarity is absent and a radial influence can manifest as the force causing bottle cell formation and the initiation of invagination. Thus cleavage is seen as a process which establishes a necessary condition for the emergence of a new phase of morphogenesis resulting from the interaction between a singularity in the surface or cortical field and the residual radial component of the cleavage field, which is described by solid harmonics.

5.6 Generation and Regeneration

There is another basic property of organisms relating the whole and the part, and this is the capacity of a part to generate or to regenerate the whole. So far, in discussing embryogenesis, the view has been developed that the fertilized egg is a whole which undergoes transformations resulting in the appearance of parts which have distinctive properties but are not, within the context of the whole organism, autonomous in the sense that atomistic theories would have them be. However, there was a time when the egg was a cell within the ovary, itself a part of the parent organism. The oocyte during its maturation develops the capacity to develop into a new whole. Such a transformation can be achieved also by parts (multicellular fragments) of hydroids such as *Hydra,* or parts (noncellular fragments) of ciliate protozoa such as *Stentor,* or a cell of a carrot, any of which can regenerate the whole organism. The capacity of plants to propagate from leaves and stems, of insects and urodeles to regenerate limbs from stumps, and of higher organisms to regenerate skin and liver, are other manifestations of this same property of parts to transform into wholes. It is evident that different organisms vary widely in their regenerative

capacities, but it is true of all organisms that from particular parts, wholes can be produced. This defines reproduction or generation, as well as regeneration, and it is one of fundamental self-organizing properties which living creatures display. What does a field description of organisms have to say about such behavior?

There is an interesting property of harmonic functions which is suggestive of precisely this capacity. If such a function is described over any part of its domain of definition (e.g., part of the surface of a sphere), then the function can be recovered uniquely over the whole domain by analytic continuation. Thus, in a particular sense, the part contains the whole. This gives us a kind of existence theorem for the generative and regenerative properties of organisms, defined as fields describable by harmonic functions. It is just this type of property that needs to be embodied in a description of biological self-organization, although the specific property of harmonic functions described here is neither necessary nor sufficient to account for the actualities of generation and regeneration in the living realm.

5.7 Structuralist Biology

It may well have become apparent to the reader before this point that the general context within which this essay has been constructed is that of contemporary Structuralism, as defined by Levi-Strauss (1968) and Piaget (1968), and developed in another, more extended analysis by Webster and Goodwin (1982). A field as used above is an example of a *structure* in that it belongs to an invariant set (the harmonic functions) defined by internal relations (those defining the field equation over a domain), each member of which is a transformation of the other (by a change of boundary values). The particular field functions which have been used in this treatment are, in a general sense, not as important as the structuralist principles which inform the analysis. For these principles emphasize the necessity to identify that which is specific to a particular area of study, such as biology, before attempting to develop a theory to "explain" the phenomena. This is why, in approaching the problem of self-organization in embryogenesis, it was necessary first to identify any empirical regularities emerging from the study of biological form which might suggest the existence of principles of organization or invariant relationships in organisms. The evidence points clearly in this direction so that organisms, and hence embryos, are indeed structures in the technical sense: i.e., entities with the defining characteristics of wholeness and self-organization, capable of undergoing transformations which preserve these deep properties while changing manifest form. This is just a more elaborated description of a view which has been clearly articulated by those who insisted that the primary problem of biology is that of organization and form, not of composition or heredity, the latter finding their place within the context of the former (cf. Russell, 1930; Needham, 1936). The use of the field concept and the more specific description of biological form in terms of harmonic functions simply makes more explicit the implications of this view. Gene products can in certain instances select specific form, such as left- or right-handed spiralling in the third cleavage planes and, consequently, in the shells of the snail, *Limnaea.* However, in general we have seen from the field description of cleavage that the constraints determining specific form arise from other organizational levels than

simple molecular composition, these being such processes as binary cleavage, animal-vegetal polarity, and a minimum "energy" condition. This emphasizes once again the primacy of organizational principles in biological process and the inadequacy of any theory based upon genes and molecules.

We may say that the role played by the genes is to specify the potential molecular composition of an organism and to define some temporal sequences in which molecular components are made. These constitute constraints which impose some limitations on the forms which organisms can assume, and in certain instances may actually determine higher-order form; but in general the relationship between "genotype" and "phenotype" is one of causal necessity, not sufficiency, since gene products act within the context of fields (electrical, visco-elastic, etc.) which generate morphology. A linguistic analogy would be that genes essentially determine the set of words out of which a text can be constructed. Words are clearly insufficient to define a text, which embodies higher-order syntactical, semantic, and; contextual constraints or rules. These are rules of organization which limit the set of allowed arrangements of the words within the text. Such organizational principles are described in this essay as field constraints, which limit the allowed range of biological forms. The generative principles of organismic morphogenesis then consist of organizational constraints or rules (laws of form) common to all organisms in the form of fields (thus defining biological universals) together with specific constraints characteristic of individual species (which then define particulars). Genes specify some of the latter but none of the former.

One of the most significant aspects of the structuralist approach is the deliberate avoidance of any *a priori* material reduction of the organism to parts such as cells, molecules, or genes. Once the problem of biological (self-) organization is clearly and explicitly described, and an appropriate description is available, then it *may* be possible to carry out a relevant material reduction. Certainly it will be possible to achieve an *abstract* reduction to laws and rules, such as those which have emerged from the analysis and description of cleavage given above. This description was itself inspired by a paper which was, in my opinion, a landmark in the discovery of general rules of morphogenesis expressed in terms which make no reference to composition or inheritance and are of a purely relational nature (French, Bryant, and Bryant, 1976). To start with the assumption that one knows what the basic parts of the organism are is to make a strategic error at the outset, which can lead one badly astray in seeking at once the most economical and the most rigorous analytical treatment of the problem. The description of organisms as fields which embody self-organizing properties and can undergo transformations preserving invariant relationships is, despite its limitations, a step towards a biological science of form and organization which relates part to whole in a manner which preserves organismic unity in the diversity of manifest morphology

Acknowledgements
I am indebted to my colleague, G.C. Webster, for the intellectual stimulus which has led to many of the ideas developed in this essay.

References

Darwin, C. 1859. *On the Origin of Species.* Reprint of the first edition, 1950. Watts & Co., London.

Driesch, H. 1929. *The Science and Philosophy of the Organism.* Black, London.

Frankel, J. 1974. Positional information in unicellular organisms. *J. Theoret. Biol,* 47: 439–81.

French, V., P.J. Bryant, and S.V. Bryant. 1976. Pattern regulation in epimorphic fields. *Science,* 193: 969–981.

Goodwin, B.C. 1980. Pattern formation and its regeneration in the protozoa. In G.W. Gooday, D. Lloyd, and A.P.J. Trinci, eds., *The Eukaryotic Microbial Cell* Symp. Soc. Gen. Microbial, 30: 377–404.

Goodwin, B.C., and N. LaCroix. 1984. A further study of the holoblastic cleavage field. *J. Theoret. Biol,* 109: 41–58.

Goodwin, B.C., and L.E.H. Trainor. 1980. A field description of the cleavage process in embryogenesis. *J. Theoret. Biol,* 85: 757–770.

Goodwin, B.C., and L.E.H. Trainor. 1982. The ontogeny and phylogeny of the pentadactyl limb. Brit. Soc. Dev. Biol., Symp. on Development and Evolution, p. 75.

Hobson, E.W. 1955. *The Theory of Spherical and Ellipsoidal Harmonics.* Chelsea Publishing Company, New York.

Levi-Strauss, C. 1968. *Structural Anthropology.* London.

Needham,J. 1936. *Order and Life.* Yale University Press.

Oosawa, F., H. Kasai, S. Hatano, and S. Asakura. 1966. Polymerization of actin and flagellin. In G.E.W. Wolstenholme and M. O'Connor, *eds., Principles of Biomolecular Organization,* pp. 273-286. Little, Brown and Co., Boston.

Piaget, J. 1968. *Structuralism.* Routledge and Kegan Paul, London.

Saunders, J.W., J.M. Cairns, and M.T. Gasseling. 1957. The role of the apical ridge of ectoderm in the differentiation of the morphological structure and inductive specificity of limb parts in the chick. *J. Morphol,* 101:57-87.

Trainor, L.E.H., and M.B. Wise. 1979. *From Physical Concept to Mathematical Structure.* University of Toronto Press.

Webster, G.C., and B.C. Goodwin. 1982. The origin of species : a structuralist view. *J. Soc. Biol. Struct.,* 5: 15-47.

Wick, G., A. Wightman, and E. Wigner. 1952. *Phys. Rev.,* 88: 101.

Wolpert, L. 1971. Positional information and pattern formation. *Curr. Top. Dev. Biol,* 6: 183-224.

Wolpert, L., and J. Lewis. 1975. Towards a theory of development. *Fed. Proc.,* 34: 14-20.

(a) 2 cells, (b) 4 cells, (c) 8 cells, (d) 16 cells, (e) 32 cells.

In (f) and (g) only, the silhouette forms one of the 8 longitudinal sections.

(a) 2 cells, (b) 4 cells, (c) 8 cells, (d) 16 cells, (e) 32 cells, (f) 64 cells, (g) 128 cells.

Case Study 17: Noise and Prediction in Biology

Perspective on *Irreproducible Results and the Breeding of Pigs (or Nondegenerate Limit Random Variables in Biology)*

Dany Spencer Adams
Research Professor
Tufts University
and
Michael Levin
Professor
Tufts University

'Bernoulli knew that life is lawful to the ensemble though chaos to the individual'

This paper is not an easy read, despite the fact that it mixes real-world examples with profound and fundamental mathematics concepts. Its applicability ranges among ecology (Ulanowicz 2007), developmental genetics (Ropers and Grimm 1977), cognitive neuroscience (Nakada 2011), the spread of disease in a population, and even cosmology. Indeed, any process where one of the variables studied is a 'non-degenerate limit random variable'. Random variable means just that. A non-degenerate limit means that chance can cause large differences in the results of replicated experiments even if they are performed identically. Flipping a fair coin is 'degenerate' because eventually, every experiment will converge on a probability of heads of 0.5. Non-degenerate means the probabilities will not converge to a single predictable number. Ignoring this can lead to false positives in all kinds of studies. Cohen gives a couple of examples, but an important point is summed up in this excerpt:

> *'I want to emphasize what almost sure convergence to a nondegenerate limit random variable looks like to people participating in a[n] experiment with this property. With increasing time, each [researcher's] experiment settles down to systematic, regular, and lawful behavior; his graphical plot of proportion [with phenotype X] as a function of time wiggles at first but smooths out gradually to a steady flat line [at a certain value]. However, if he repeats the experiment or gets a friend to do so under identical conditions, … the curve of the replicate experiment levels out [at a value that] seems to bear no relation to the original. It is only after a change in the level of analysis—only after considering an ensemble of replications—that regularity and simplicity reappear.'*

In other words, if you are dealing with a non-degenerate limit random variable, replicates of experiments (each of which comprises many samples) may be performed identically, but there *will* nonetheless be differences among results. To learn about the population, not just about sets of samples, requires examination of the results from 'ensembles' of replicated experiments.

More broadly, the paper explores the highly counter-intuitive dynamics of recursive processes, so prevalent in biology, where events are modified by the outcomes of prior events. Under those conditions, small initial differences can be

amplified toward distinct outcomes. This not only makes prediction difficult, but can mislead one toward inferring the presence of deterministic causes where the role of chance is sufficient to explain strongly divergent results from identical experimental setups. Though not often cited, the concepts in this paper (and the even earlier discussion of Polya's Urn (Eggenberger and Polya 1923)) are fundamental to the understanding of chance, causation, and prediction in biology.

Cited:

Eggenberger F and Polya G 1923 Uber die statistik verketteter vorgange *Z Angew. Math. Mech.* **3** 279–89

Nakada T 2011 Brain science of the mind *Soc. Sci. Inform.* **50** 25–38

Ropers H H and Grimm T 1977 Variable composition of X chromosomal mosaics: due to asynchronous cell division during early embryogenesis? *Hum. Genet.* **39** 213–5

Ulanowicz R E 2007 Emergence, Naturally! *Zygon* **42** 945–60

Irreproducible Results and the Breeding of Pigs (or Nondegenerate Limit Random Variables in Biology)

Joel E. Cohen

Many people believe that a coin in ordinary currency will come up heads nearly half the time it is tossed. Few people have reported a systematic experimental test of *that* belief.

During World War II an English statistician, J. E. Kerrich, was in Denmark when the Germans overran it. Interned under benevolent Danish supervision, he performed and recorded (Kerrich 1946) 10,000 spins of an ordinary coin. The proportions of trials which came up heads after ten, a hundred, a thousand, and ten thousand trials were, respectively, 0.400, 0.440, 0.502, and 0.507. If the ten thousand trials are broken into ten blocks of a thousand trials each, then the proportions of heads after each of the ten blocks were 0.502, 0.511, 0.497. 0.519, 0.504, 0.47 6, 0.507, 0.5I8, 0.504, and 0.529.

Viewed as a single long series, the data show that the proportion of heads tended toward and remained near one-half as the number of trials (tosses or spins) increased. Viewed as ten shorter series, the data suggest that the proportions of heads in independent experiments under the same conditions tended toward a single common value.

In 1713, Bernoulli constructed a mathematical idealization of the coin-tossing experiment as a sequence of independent trials each with a fixed probabilily p of coming up heads. Here p is some fraction near one-half. Imagine a very large number of copies of Kerrich all tossing copies of the same coin under the same conditions in perfect synchrony, but with the outcome of each coin toss independent of every other outcome. Bernoulli showed that, as the number of tosses increases, the proportion of all the copies of Kerrich for each of whom the fraction of his trials coming up heads differs from p by less than some arbitrarily small fixed amount approaches 100%. Mathematicians call this phenomenon convergence in probability to the constant limit p.

Two centuries later, in 1909, Borel proved that the same imaginary situation is even more lawful than Bernoulli had supposed. Bernoulli's result does not rule out the possibility that the proportion of heads in the trials of a particular copy of Kerrich could continue indefinitely to wander away from p by at least some fixed nonzero amount. Borel ruled out this possibility: for 100% of the copies of Kerrich, as the number of each man's tosses increases, the proportion of his trials coming up heads must approach and remain arbitrarily near the value p. Mathematicians call this phenomenon convergence with probability 1 or almost sure convergence to the constant limit p (see Loève 1963, pp. 14 and 19 for short proofs). Few people find these results a shock to their intuition.

The author is with Rockefeller University, New York, NY 10021.

ONE GOOD URN

Now consider an equally simple experiment. Suppose a very large box (whose capacity can be extended indefinitely by adjunction of similar boxes) initially contains one green ball and one blue ball. Choose one ball at random, look at its color, replace the ball in the box, and add to the box another ball of the same color as the one chosen. At each successive point in time, say once every second, choose one ball at random and then repeat exactly the above. The precise meaning of "at random" is that if there are n balls in the box when a drawing is made, each ball has an equal chance $1/n$ of being drawn.

The proportion of green balls in the box is the number of green balls divided by the total number of balls, whether blue or green. What will happen to the proportion of green balls as time increases?

Before reading further, please make a serious effort to guess. You have three guesses. When I proposed this problem to a very august mathematical ecologist in the course of a country march, he gave up after two wrong guesses. When I first heard the answer myself, I was astonished both by the general phenomenon it exemplifies and by the particular details. On reflection, I think the general phenomenon permeates population biology. My purpose here is to describe the phenomenon, give some biological examples of it, and suggest its consequences for the interpretation of biological data.

The experiment just described is a special case of what is known as "Polya's urn scheme." Eggenberger and Polya (1923) introduced the scheme in 1923 to model the spread of infection in a population. David Blackwell and David Kendall (1964) studied another generalization of this experiment and even mentioned its implications for stochastic population growth. But an overgrowth of related mathematical results obscured their message for biologists.

So, suppose a single Kerrich performs the experiment with blue and green balls. As time goes on, the proportion of green balls will converge to some limit p. (This fact alone is not obvious.) As in the Bernoulli model, "converge" means that if you pick some nonzero tolerance interval around p, then there is some time at which the proportion of green balls will be in that tolerance interval and after which it will never leave it; the proportion approaches and remains near $\boldsymbol{p}$.

But what is p? For the first copy of Kerrich, call him Kerrich_1, all one can say is that his value of p, call it p_1, lies between 0 and 1 inclusive The chance that his p_1 is exactly equal to any particular fixed p between 0 and 1 is zero! However, the chance that his p_1 falls between 0.2 and 0.3 inclusive is exactly $0.3-0.2 = 0.1$.

At the next desk, Kerrich_2 is finding that the proportion_2 of green balls in his box_2 is getting and remaining closer and closer to a fixed number p_2. But whereas Kerrich_1's proportion seems to be approaching $p_1 = 0.2435871\ldots$, his proportion_2 is approaching $p_2 = 0.9342265\ldots$ And on his other side, Kerrich_3's proportion is approaching $p_3 = 0.59943312\ldots$ Each man's proportion of green balls converges to a limit, which is constant for each particular man but which varies from one man to another, even though all- are performing exactly the same experiment. In this case, the limiting proportion p is uniformly distributed over the interval from 0 to 1: that

is, the chance that Kerrich$_{17}$'s limit p_{17} falls between a and b, where $0 \leqslant a \leqslant b \leqslant 1$, is $b - a$.

Blackwell and Kendall (1964) proved that if the box starts out with one ball of each of k different colors, where k may exceed 2, then the limiting distribution of proportions of each of the k colors is uniform over the set of all possible ways of dividing 100% into k proportions.

The behavior of this hypothetical experiment exemplifies what mathematicians call almost sure convergence to a nondegenerate limit random variable. "Nondegenerate" means that the limiting value of the proportion of green balls is not restricted to a single point. In Bernoulli's model of the real experiment which the real Kerrich performed, the limit random variable is degenerate because every such Kerrich would (in theory) have obtained the same limiting proportion p of heads.

This hypothetical experiment behaves identically to an apparently quite different experiment. Suppose each Kerrich has a box with one green and one blue ball. He receives a coin; one side is green, one blue. Though the coin looks fair, the real probability p that the coin will come up green is distributed uniformly between 0 and 1. For any given coin, p is constant in time. No man has any reason to suspect that his coin differs from any of the others; in particular, he does not know his coin's value of p. Once a second, each man flips his coin and adds to his box a ball of the color indicated. Then (since this experiment is just Bernoulli's model and Borel's theorem applies) each man's proportion of green balls converges with probability 1 to his coin's p.

Here each man's limiting proportion p is assigned first. The color of the next ball is chosen by an independent trial with probability p of green. In the previous hypothetical experiment, each new ball's color is determined by random choice among the colors which have occurred so far. To an observer of the balls deposited in the boxes, the two experiments are indistinguishable[3]

Before proceeding to biological examples. I want to emphasize what almost sure convergence to a nondegenerate limit random variable looks like to people participating in an experiment with this property. With increasing time, each man's experiment settles down to systematic, regular, and lawful behavior; his graphical plot of proportion green as a function of time wiggles at first but smooths out gradually to a steady flat line. However, if he repeats the experiment or gets a friend to do so under identical conditions, where the curve of the replicate experiment levels out seems to bear no relation to the original.

It is only after a change in the level of analysis–only after considering an ensemble of replications–that regularity and simplicity reappear. It is the law of the limit random variable that is simple.

In retrospect, Bernoulli himself made just such a change in the level of analysis. If each copy of Kerrich were to toss a coin just once, then Kerrich1 might get heads, Kerrich2 tails, Kerrich3 again heads, and so on without apparent pattern. As the size of the ensemble of copies of Kerrich increases, however, the proportion of copies

[3] Violet Cane (197 3) has discovered an equally surprising, and closely connected, observational equivalence of models for negative binomially distributed counts, such as accident statistics.

whose single trial results in heads approaches the limit p near one-half. Already Bernoulli knew that life is lawful to the ensemble though chaos to the individual.

BACK ON THE FARM

Now suppose that a breeding stock on a pig farm is maintained by mating a boar and a sow each generation. One male and one female from the offspring are chosen to mate in the next generation. Suppose there is a single gene locus at which, in the initial generation, both parents are heterozygotes. For example, each has genotype Aa. Since each off-spring receives one allele chosen at random from each of its parents, there is positive probability that both offspring will have the genotype aa. If this happens, all future offspring will have the same genotype at that locus. There is an equal positive probability that both offspring will have the same genotype AA, with the same consequence.

Sooner or later both offspring must become homozygous for the same allele, and geneticists have calculated the rate of approach to homozygosity under the regime of inbreeding just described. The offspring of a particular pair of heterozygous parents will fixate on the genotype aa with probability one-half and on the genotype AA with probability one-half.

Aside from their good looks and intelligence, pigs are bred for characteristics of commercial interest such as quantity of edible meat. These quantitative characters are believed to be controlled by the additive effects of genes at several loci. Suppose, for the sake of illustration, that weight is controlled by five independently assorting loci with alleles A, a; B, b; and so on up to E, e. Let homozygosity for the capital letter at a locus correspond to an increase in one kilogram over the heterozygote and homozygosity for the small letter at a locus correspond to a decrease in one kilogram below the heterozygote.

If a breeding line is started with parents both of genotype AaBbCcDdEe, then eventually the descendants in that line are certain to drift to homozygosity, at each locus the same for both male and female. The weight of pairs in successive generations will cease fluctuating eventually, all else being equal, and will be the same for both members of the pair. Their weight at fixation will be 5 kilograms above those of their initial ancestors if all five loci fixate at capital letters, 3 kilograms above if four of the five loci fixate at capitals. 1 kilogram above if three of the five fixate at capitals, or symmetrically below the weights of their ancestors. The weight at fixation of another line of descent might differ. As the size of an ensemble of lines of descent increases, the proportions of lines at each weight approach the probabilities calculated from a binomial distribution with parameters 5 and 1/2 (roughly, a bell-shaped histogram with its highest values symmetrically placed on either side of the ancestral weight).

When a selective breeding program uses a finite stock of pigs (and infinite numbers of pigs have not yet been observed), this underlying drift, due to random sampling of genes, sets limits to what selection can accomplish. Moreover, drift to a nondegenerate limit random variable sets different limits in different replications of an identical breeding program. As Robertson (1960, p. 244) observed: "If $u(q)$ [the chance of fixation of an allele whose frequency at the beginning of a breeding

program is q 1 is very different from unity for many genes, we will notice that replicate lines from the same initial population will be very different in the limit they reach." In our example, $q = 1/2$ and $u(q) = 1/2$ for all five loci, so Robertson's warning applies.

A failure to recognize the nondegeneracy of the limit random variable to which polygenic characters drift has practical consequences. Hill (1971, p. 294) points out that some authors estimate realized heritability in a single selection program "by fitting a linear regression to cumulative response and cumulative selection differential each generation. But with genetic sampling (drift) the variance of the population mean increases each generation, and these means become correlated. In standard regression analysis the observations are assumed to have equal variance and be uncorrelated, so that the estimates of variance of realised heritability obtained by. . . . using standard regression techniques are biassed downwards. In other words, the observed variance among heritability estimates from a replicated experiment would exceed the variance predicted from a single replicate." Hill gives an explicit quantitative analysis of what nondegenerate drift does and what to do about it in an important series of papers (most recently, 1974).

While geneticists have long known of genetic drift and have recently assayed its practical impact on breeding programs, other areas of population biology seem to have remained in bliss. Suppose two bacteria, say a wild type and a mutant, are distinguishable by some marker but are absolutely identical with respect to growth in a particular culture, which is sufficiently favorable to growth that no deaths occur. After a while one or the other of the bacteria will divide, giving three bacteria. Then, one of those three will divide, each one being equally likely, and so on. If we ignore the interval between divisions and advance an artificial clock by one unit at each division in the culture, we obtain exactly the Polya urn model. If we identify the wild type with green balls, and the mutant with blue, then after a long time, since the proportion of green balls converges to a limit, so will the proportion of wild type bacteria in the culture, and to the same limit. The chance that this limit is exactly one-half is zero. lf the limit is p, the culture would behave as if each new bacterium added were wild type with probability p. Blackwell and Kendall (1964, p. 295) state succinctly: "This might lead the incautious observer to attribute a real difference to the … clones in respect of their growth mechanism, although in fact they are in all ways identical." The same phenomenon might lead incautious observers to infer that a genetic change affecting growth had occurred if they attempted to replicate the experiment and found, as they must, a different limiting proportion of the wild type.

Similarly, suppose that individuals of a growing population fall into one of k age categories, where k may exceed 2. Under certain assumptions (Athreya and Ney 1972, p. 206), which may even be defensible in some real situations, the proportions in each class will approach proportions which depend only on the fertility and mortality, but not on the initial numbers of individuals, of each age class. Moreover, the population will (with positive probability) eventually grow exponentially at a rate which also depends only on the fertility and mortality of each age class. If total population size is plotted on a logarithmically scaled ordinate against time on the abscissa, the graph will eventually fall along a straight line. The point at which this

straight line intersects the time axis is where a deterministically growing population with the same growth rate would have had to begin growing exponentially in order to fall into step alongside the stochastically modeled population. Call this point the lag time. It is a nondegenerate random variable. Though the laws of growth are the same, the lag times, or times to apparent exponential take-off of growth, of initially identical populations obeying this stochastic model are different.

No deterministic interpretation of such differences in limiting proportions or lag times could possibly be right, though the differences are real enough. The variation in an ensemble of replicates must become the object of study when the limit random variable of an individual replicate is nondegenerate. Luria and Delbrück (1943) practiced this precept in their classic experimental proof that phage-resistant mutants arise randomly.

But population biologists who study macroscopic populations seem less inclined to this view of nature. Here are a few heretical possibilities Is it possible that differences in successional changes and in so-called climax state in apparently similar habitats are not to be explained as due to any causal difference between the habitats, but should be interpreted as variation in an ensemble of such habitats? Is it possible that the differences in species composition of apparently similar islands result from the operation of identical forces which produce regularity only in an ensemble of islands? Could observations that one animal population cycles with a period of 4 years and another with a period of 3 or 13 or 17 years become intelligible if the ensemble of periods of cycling animal populations were examined? Could the differences in the sizes of prides of lions or in the social organization of troops of Japanese' macaques reflect the inherently but lawfully variable outcome of identical underlying stochastic forces, rather than deterministic ecological differences?

In behavior, for example, is it possible that some of the significant differences among mother-child interactions, which are obvious by the time a child reaches five years of age, are due neither to inherent differences among individuals nor to environmental differences, but to sequentially dependent random forces applying equally to all mother-infant pairs? (Which hand the mother first holds the infant with is random, perhaps, but that choice affects the skill with which she performs tasks with her remaining hand, which affects the infant's response, wh1ch affects.…) Could nondegenerate stochastic limits provide useful models of what students of plant development (Evans 1972) call "ontogenetic drift"?

The possibilities I raise will leave cold or enrage people who believe they *know* that deterministic factors explain some of the differences I cite. They may well be right in part. All I suggest is that there may be variation which deterministic factors do not usefully explain and that the possibility of understanding phenomena is preserved by redirecting attention to a lawful-looking ensemble.

CHAOS AND COSMOLOGY

May (1974, p. 645) has emphasized the ecological interpretation of the mathematical fact that very simple deterministic difference equations can have astonishingly messy trajectories, including "cycles of any period, or even totally aperiodic but bounded … fluctuations." The recent, still unpublished work of several people shows that

many (though not necessarily all) difference equations studied by ecologists can act so weirdly. Implicit is the suggestion that the apparent variability of population fluctuations may represent the working of a simple deterministic mechanism. The behavior predicted by this mechanism is so sensitive to the values of the parameters in at least some ranges that it will probably be necessary to compare observations with a probabilistic approximation. Thus, the apparent variability of population fluctuations, for example, is interpreted at two levels in the models May and others consider: first, in the complexity of the trajectories predicted with fixed parameters (including initial values); and second, in the impossibility of estimating exactly, and the likely actual fluctuation of, the parameter values. These models do not attempt to account for uncertainty or fluctuations in parameters but assume, at the kernel of phenomena, a simple determinism.

It seems impossible to reject with any data an affirmation of faith that a deterministic mechanism could supply sufficient apparent variability to describe a real population whose parameters were known and constant. The preceding biological examples, and others which could be cited, suggest an alternate view: At least some biological processes incorporate stochastic elements that can cause long-term behavior which appears lawful only in an ensemble of replicates. The empirical program suggested by this view is to examine such ensembles.

Worn exclusively, the deterministic glasses of Laplace and the stochastic glasses of Charles Sanders Peirce give equally roseate views of the world. In the interest of fair advertising, I have to admit that the strategy of moving from the individual to the ensemble to find order in variability will not always work. There are stochastic processes which approach a limit (any kind of limit, degenerate or not) only with a probability 0. Some misanthropes claim experience is like that, too: Some parts of nature simply change more slowly than others, they say, and those parts that change slowly compared to the time scale we are interested in serve us as points of reference, or limits built on sand. Such misanthropes may be right.

Having speculated thus far, let me raise and answer a metaphysical question prompted by Polya's urn which would, I hope, have amused Peirce as it amuses me on dark nights. If you and I had been- born in another universe which had started from exactly the same initial conditions as our present one and which had been subject to the same dynamics, would we necessarily infer the same Jaws of nature as we (in the collective sense of civilized thought) infer for this universe? I take the existence of genetic drift on pig farms as establishing a stochastic element in the dynamics of the universe, and therefore have no guarantee that the apparent lawfulness in this copy of the universe would take the same form in any other. The order of this universe may be an irreproducible result.[4]

ACKNOWLEDGMENTS

This paper is based in part on a talk given in the Provost's Lodge, King's College, Cambridge, 25 April 1975. For extraordinary hospitality during 1974-75, I thank

[4] Anne Whittaker points out that Ray Bradbury has dramatized possibility (see Bradbury 1962).

members of King's College Cambridge and the University of Cambridge Statistical Laboratory. I thank Robin Sibson for excellent teaching; and John Hajnal, David Kendall, and the referees for helpful comments. This work was supported by King's College Research Centre and the U.S. National Science Foundation.

REFERENCES CITED

Athreya, K. B., and P. E. Ney. 1972. *Branching Processes.* Springer, New York.

Blackwell, D. and D. G. Kendall. 1964. The Martin boundary for Polya's urn scheme and an application to stochastic population growth. *J. Appl. Probab.* 1(2): 284-296.

Bradbury, R. 1962. A sound of thunder. Reprinted in *R Is for Rocket.* Doubleday, Garden City, New York, 195 2.

Cane, V. R. 1973. The concept of accident proneness. *Académie Bulgare a$$ Sciences Bulletin de l'nstitut de Mathématiques* 15: 183-188.

Eggenberger, F., and G. Polya. 1923 Uber die Statistik verketteter Vorgänge. *Zeuangew. Math. Mech.* 3: 279-289.

Evans, G. C. 1972. *The Quantitative Analysis of Plant Growth.* Blackwell Scientific, Oxford.

Hill, W. G. 1971. Design and efficiency of selection experiments for estimating genetic parameters. *Biometrics* 27: 293-311.

———. 1974. Variability of response to selection in genetic experiments. *Biometrics* 30: 363-366.

Kerrich, J. E. 1946. *An Experimental Introduction to the Theory of Probability.* Einar Munskgaard, Copenhagen.

Loève, M. 196 3. *Probability Theory,* 3rd ed. D. Van Nostrand, Princeton.

Luria, S. E., and M. Delbrück. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28: 491-511.

May, R. M. 1974. Biological populations with nonoverlapping generations: stable points, stable cycles, and chaos. *Science* 186: 645-647.

Robertson, A. 1960. A theory of limits in artificial selection. *Proc. R. Soc. Land.* ***B*** *Bioi. Sci.* 153: 234-249.