Information Transport and Evolutionary Dynamics

Marc Harper

NIMBioS Workshop

April 2015

Thanks to

- NIMBioS
- John Baez
- All the participants

Motivation

Theodosius Dobzhansky (1973)

Nothing in biology makes sense except in the light of evolution.

Richard Dawkins, *The Blind Watchmaker* (1987)

The theory of evolution by cumulative natural selection is the only theory we know of that is in principle capable of explaining the existence of organized complexity.

Motivation

Donald T. Campbell, *Evolutionary Epistemology* (1974)

A blind-variation-and-selective-retention process is fundamental to all inductive achievements, to all genuine increases in knowledge, to all increases in the fit of system to environment.

Ronald Fisher, *The Design of Experiments* (1935)

Inductive inference is the only process known to us by which essentially new knowledge comes into the world.

Universal Darwinism

Richard Dawkins proposed a theory of evolutionary processes called Universal Darwinism (*The Selfish Gene*, 1976), later developed further by Daniel Dennett (*Darwin's Dangerous Idea*, 1995) and others.

An **evolutionary process** consists of

- Replicating entities that have heritable traits
- Variation of the traits and/or entities
- Selection of variants favoring those more fit to their environment

See also Donald T. Campbell's BVSR: Blind Variation and Selective Retention (1960s)

Replication

What is replication? The proliferation of some static or dynamic pattern (*Lila*, Robert Pirsig, 1991). A **replicator** is something that replicates:

- Biological organisms
- Cells
- Some organelles such as mitochondria and chloroplasts

Hummert et al. *Evolutionary game theory: cells as players.* Molecular BioSystems (2014)

Replication

Molecular replicators:

- Genes, transposons, bacterial plasmids
- Self-replicating RNA strands (Spiegelman's monster)
- Viruses, RNA viruses (naked RNA strands)
- Prions, self-replicating proteins e.g. Bovine spongiform encephalopathy

Bohl et al. *Evolutionary game theory: molecules as players.* Molecular BioSystems (2014)

Replication

James Gleick, *The Information: A History, a Theory, a Flood* (2012)

Evolution itself embodies an ongoing exchange of information between organism and environment .... The gene has its cultural analog, too: the meme. In cultural evolution, a meme is a replicator and propagator an idea, a fashion, a chain letter, or a conspiracy theory. On a bad day, a meme is a virus.

- Meme: an idea, behavior, or style that spreads from person to person within a culture (Dawkins)
- Words, sounds, phrases, songs, the alphabet
- Proverbs: "Those who live in glass houses shouldn't throw stones" – Ancient Egypt
- Language itself

Tree of Languages

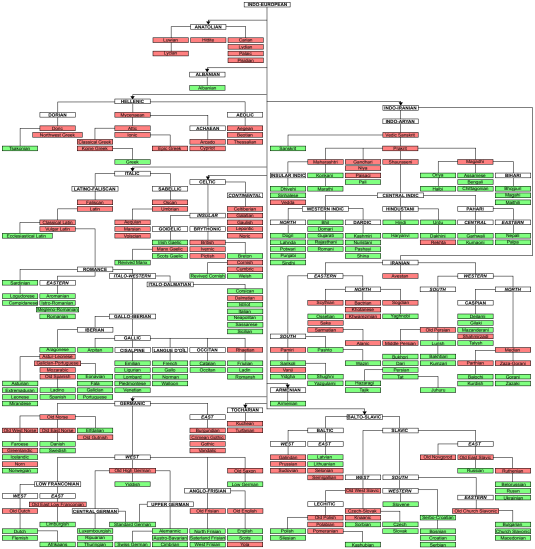


Image source: https://en.wikipedia.org/wiki/Proto-Indo-European_language

Phylogenetically Determined

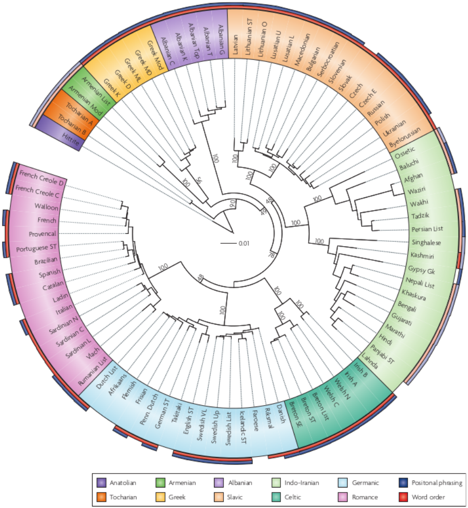


Image source: Mark Pagel, *Human Language as a Culturally Transmitted Replicator* (2009)

Shannon's Information Theory

John von Neumann, speaking to Claude Shannon (circa 1949)

You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage.

Information Theory

Shannon's information theory is concerned with transmission of information through possibly noisy channels

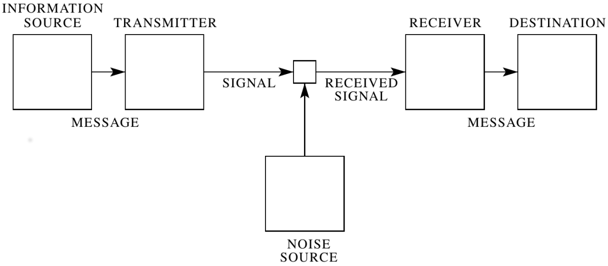


Fig. 1—Schematic diagram of a general communication system.

Figure source: *A Mathematical Theory of Communication*, Claude Shannon, 1948

Information Theory

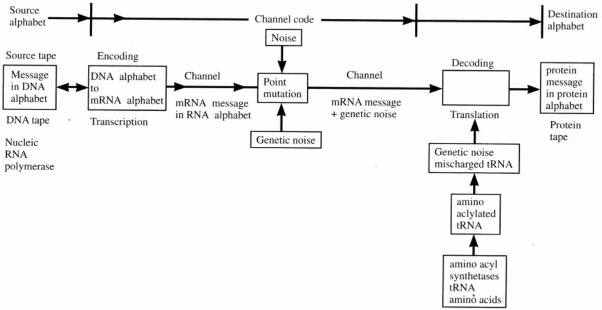


Figure 5.2. The transmission of genetic messages from the DNA tape to the protein tape as conceived in molecular biology. Genetic noise occurs in all stages but is lumped in the figure to fix the idea.

Figure source: Hubert Yockey, *Information Theory, Evolution, and the Origin of Life*, 2005

Replication and Information Theory

- Replication has a very natural information-theoretic interpretation – information transport across noisy channels
- In reality all channels are noisy – e.g. thermodynamic fluctuations for replicating molecules
- Accordingly, replication necessarily has variation

Evolution

Let's recast Universal Darwinism

- Natural proliferation of patterns – transmission through noisy channels
- Natural diversification of patterns – acquired variations during transmission
- What about selection?
- Replicators proliferate at different rates depending on environment and other factors

Selection

- In evolutionary dynamics it is common to specify a **fitness landscape** rather than to attempt to correctly model growth rates
- A fitness landscape is analogous to a potential in physics (but we seek to maximize rather than minimize)
- A key contribution of EGT is that replicators themselves are part of the environment
- Often we look at fitness landscapes of the form $f(x) = Ax$ where A is a game matrix

The Replicator Equation

A popular model of populations under the influence of natural selection is the *replicator equation*.

Consider population composed of n types of replicators (e.g. phenotypes, genotypes, species) $T_1, \ldots, T_n$ with proportions $x_1, \ldots, x_n$. Let the fitness landscape be a vector-valued function $f(x)$ where $f_i(x)$ is the fitness of type T_i. Then:

relative rate of change of type T_i = fitness of type T_i − mean fitness

$$\frac{1}{x_i}\frac{dx_i}{dt} = f_i(x) - x \cdot f(x)$$

$$\boxed{\frac{dx_i}{dt} = x_i\,(f_i(x) - x \cdot f(x))}$$

The Discrete Replicator Equation

There is also a discrete time model

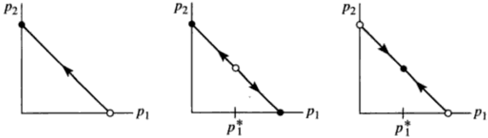
$$x_i' = \frac{x_i f_i(x)}{x \cdot f(x)} \qquad P(H_i|E) = \frac{P(H_i)P(E|Hi)}{P(E)}$$

Essentially the same as Bayesian inference, first observed (?) by C. Shalizi circa 2007-2008

Replicator Phase Plots – Two Types

$$\frac{dx_i}{dt} = x_i \left(f_i(x) - x \cdot f(x) \right)$$

24 Chapter 2 Symmetric Normal Form Games



(a) Prisoner's Dilemma (b) Coordination (c) Hawk-Dove

Figure 2.2.1
Phase portrait of the replicator dynamic for two-strategy games. Trajectories lie on the line $p_1 + p_2 = 1$. Circles indicate rest points of the dynamic (solid are stable and empty unstable) while arrows indicate increasing t.

Source: Ross Cressman, *Evolutionary Dynamics and Extensive Form Games*, MIT Press (2003)

Replicator Phase Plots – Three Types

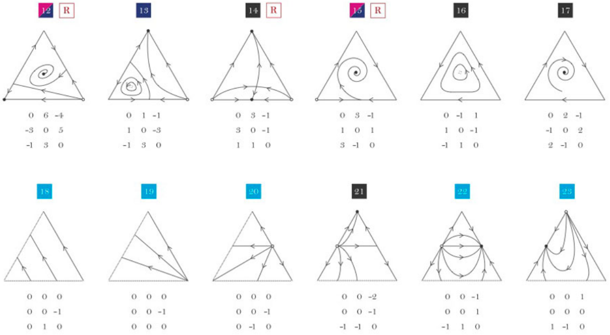
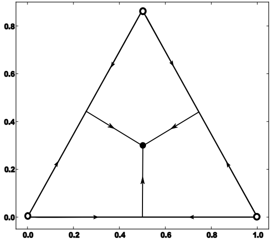
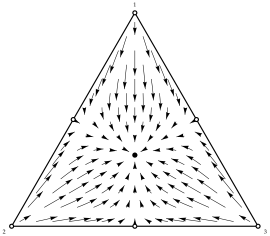


Image source: Aydin Mohseni, http://www.amohseni.net/
Classification: Bomze, I. *Lotka-Volterra equation and replicator dynamics: a two-dimensional classification.* Biological cybernetics (1983)

Replicator Phase Plots – Three Types

$$\text{Let } A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \qquad f(x) = Ax$$





Right image made made with Bill Sandholm's *Dynamo*.

Evolutionary Stability

$$\frac{1}{x_i}\frac{dx_i}{dt} = (f_i(x) - x \cdot f(x))$$

$$D(y,x) = \sum_i y_i \log y_i - \sum_i y_i \log x_i$$

Evolutionarily Stable State if for a neighborhood of e:

$$e \cdot f(x) > x \cdot f(x)$$

Theorem: Given an interior evolutionarily stable state e, it is easy to show that $D(e,x)$ is a local Lyapunov function for the replicator equation:

$$\frac{d}{dt}D(e,x) = -\sum_i e_i \frac{1}{x_i}\frac{dx_i}{dt} = x \cdot f(x) - e \cdot f(x) < 0$$

Hofbauer 1978, Akin and Losert 1984, Bomze 1991, Amari 1995;
ESS: John Maynard Smith and George Price ˜1972

The Replicator Equation, Information-Geometrically

The second way is through information geometry. We can show that the replicator equation is a gradient flow with respect to a local information measure (Kimura's Maximal principle).

$$\frac{dx}{dt} = \nabla V(x)$$

Riemannian Geometry

In Riemannian geometry we use a **metric** to specify the geometry of a space. The angle between vectors depends on the dot product:

$$\langle a, b \rangle = a \cdot b = |a||b| \cos(\theta)$$

where $a \cdot b = \sum_i a_i b_i$

By altering the dot product we can change the angles between tangent vectors and impart curvature to a space.

Riemannian Geometry

We do this by introducing a matrix-valued function g (called the **metric**) that changes the dot product at every point in a space. At the point x,

$$\langle a, b \rangle_x = \sum_{ij} a_i \left[g(x) \right]_{ij} b_j$$

For standard Euclidean space, the matrix is always the identity at every point.

Information Geometry

In information geometry, the *points* of our space are probability distributions and the metric is called the **Fisher information metric**.
If we consider the space of discrete probability distributions (the simplex), i.e. all points of the form $x = (x_1, ..., x_n)$ with x_i real-valued, $0 < x_i < 1$ and $x_1 + \cdots + x_n$, the **Fisher information metric** takes the form

$$g_{ij}(x) = \frac{1}{x_i}\delta_{ij},$$

where δ_{ij} is the Kronecker delta.
Then the dot product becomes

$$\langle a, b \rangle_x = \sum_i \frac{a_i b_i}{x_i}$$

Information Geometry

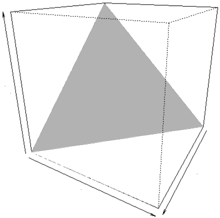
So this is a weird (looking) metric by most accounts, but it turns out that we can transform this space into a sphere! Letting $y_i = 2\sqrt{x_i}$, we have that

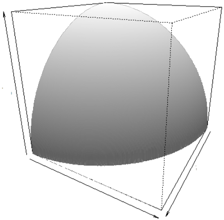
$$1 = x_1 + \cdots + x_n = y_1^2/4 + \cdots + y_n^2/4,$$

$$2^2 = y_1^2 + \cdots + y_n^2$$

The metric transforms from the Fisher information metric to the Euclidean metric on the surface of the sphere or radius 2.

Information Geometry





Slices (x_3 constant) map to quarter circles ($y_3 = 2\sqrt{x_3}$ constant):

$$1 - x_3 = x_1 + x_2 \longrightarrow 4 - y_3^2 = y_1^2 + y_2^2$$

Replicator Equation – Information Geometry

- In information geometry, the gradient flow on the Fisher information geometry is called the *natural gradient*
- In evolutionary dynamics it's called the replicator equation on the Shahshahani geometry
- Precisely, if the fitness landscape is a Euclidean gradient (∇V for some function $V : \Delta^n \to \mathbb{R}$), then the right-hand side of the replicator equation is also a gradient with respect to the Fisher/Shahshahani metric.
- Let $V = \frac{1}{2} x \cdot Ax$ be the mean-fitness, with $A = A^T$ a symmetric matrix. Then $\nabla V = Ax = f(x)$, and the (information) gradient is

$$\nabla_S V = g^{-1}\left((\nabla V)_i - x \cdot \nabla V\right) = x_i(f_i(x) - x \cdot f(x)).$$

Replicator Equation – Information Geometry

These results have led some to describe the replicator equation as an **information transport** equation (Jerome Harms, *Information and Meaning in Evolutionary Processes* (2004)).

A Taylor expansion of the relative entropy has the Fisher matrix as the Hessian (second derivative):

$$\begin{aligned} D(y,x) &= \sum_i y_i \log y_i - \sum_i y_i \log x_i \\ &= \frac{1}{2}(x-y)^T g(x)(x-y) + \cdots \end{aligned}$$

Fisher's Fundamental Theorem

We also get a nice statement of FFT, again with mean-fitness $V = \frac{1}{2}x \cdot Ax$, $A = A^T$:

$$\boxed{\frac{dV}{dt} = \mathrm{Var}_x\left[f(x)\right] = \sum_i x_i \left(f_i(x) - x \cdot f(x)\right)^2}$$

$$\begin{aligned}
\frac{dV}{dt} &= \frac{1}{2}\frac{d}{dt}\left(x \cdot Ax\right) = \frac{1}{2}\frac{dx}{dt} \cdot Ax + \frac{1}{2}x \cdot A\frac{dx}{dt} \\
&= \frac{dx}{dt} \cdot Ax = \sum_i x_i(f_i(x) - \bar{f}(x))f_i(x) \\
&= \sum_i x_i(f_i(x) - \bar{f}(x))(f_i(x) - \bar{f}(x)) = \mathrm{Var}_x\left[f(x)\right]
\end{aligned}$$

Generalizations of this Story

- Starting from any generalized relative entropy, one can get a generalized replicator equation, a Lyapunov theorem, etc.
- Additional game dynamics – best reply, logit, BvNN, etc.
- Discrete and other time scales (time-scale calculus)
- Other probability distributions: *Gradient Systems in View of Information Geometry*, Fujiwara and Amari 1995

Finite Populations

The replicator equation is a nice model, but it's unrealistic in many ways:

- Infinite population size, no drift
- No mutation or stochasticity

It turns out that information theory can be used to understand finite populations modeled as Markov processes in a very similar way.

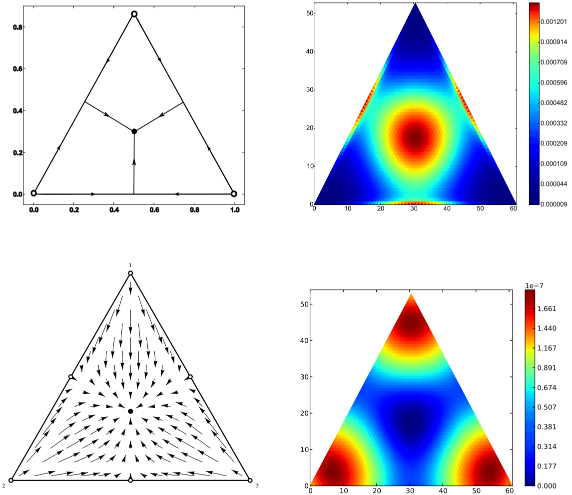
Finite Populations

By adding mutation to e.g. the Moran process we obtain processes without stationary states that have similar "equilibria" that minimize an information-theoretic quantity.
We do this by looking at the *expected next state* of a process

$$E(x) = \sum_{x \to y} y T_x^y$$

Then we have that the maxima and minima of the stationary distribution occur when $E(x) = x$, and when $D(E(x), x)$ is minimized.

Finite Population Example



Top Right: Stationary Distribution — Bottom right: $D(E(x), x)$

Thanks!

Questions?

The Moran Process with Mutation

The distribution of fitness proportionate selection probabilities is given by $p(\bar{a}) = M(\bar{a})\bar{\varphi}(\bar{a})$ where $\varphi_i(\bar{a}) = \bar{a}_i f_i(\bar{a})$; explicitly, the i-th component is

$$p_i(\bar{a}) = \frac{\sum_{k=1}^{n} \varphi_k(\bar{a}) M_{ki}}{\sum_{k=1}^{n} \varphi_k(\bar{a})}$$

The transition probabilities are:

$$\begin{aligned} T_a^{a+i_{j,k}} &= p_j(\bar{a})\bar{a}_k \qquad \text{for } j \neq k \\ T_a^a &= 1 - \sum_{b \text{ adj } a, b \neq a} T_a^b \end{aligned} \tag{1}$$

Stationary Stability

- It's easy to show that for the Moran process

$$E(\bar{a}) = \frac{1}{N} \sum_{b \text{ adj a}} bT_a^b = \frac{N-1}{N}\bar{a} + \frac{1}{N}p(\bar{a})$$

- For the Wright-Fisher process $E(\bar{a}) = p(\bar{a})$ (multinomial expectation)
- For both processes, $E(\bar{a}) = \bar{a}$ iff $\bar{a} = p(\bar{a})$